JOURNAL OF
INFORMATION SYSTEMS APPLIED RESEARCH

In this issue:

The **Journal of Information Systems Applied Research** (JISAR) is a double-blind peer-reviewed academic journal published by **ISCAP,** Information Systems and Computing Academic Professionals. Publishing frequency is currently semi-annually. The first date of publication was December 1, 2008.

JISAR is published online (http://jisar.org) in connection with CONISAR, the Conference on Information Systems Applied Research, which is also double-blind peer reviewed. Our sister publication, the Proceedings of CONISAR, features all papers, panels, workshops, and presentations from the conference. (http://conisar.org)

The journal acceptance review process involves a minimum of three double-blind peer reviews, where both the reviewer is not aware of the identities of the authors and the authors are not aware of the identities of the reviewers. The initial reviews happen before the conference. At that point papers are divided into award papers (top 15%), other journal papers (top 30%), unsettled papers, and non-journal papers. The unsettled papers are subjected to a second round of blind peer review to establish whether they will be accepted to the journal or not. Those papers that are deemed of sufficient quality are accepted for publication in the JISAR journal. Currently the target acceptance rate for the journal is about 40%.

Questions should be addressed to the editor at editor@jisar.org or the publisher at publisher@jisar.org. Special thanks to members of AITP-EDSIG who perform the editorial and review processes for JISAR.

JOURNAL OF
INFORMATION SYSTEMS APPLIED RESEARCH

Editors

Scott Hunsinger
Senior Editor
Appalachian State University

Thomas Janicki
Publisher
University of North Carolina Wilmington

2017 JISAR Editorial Board

Identifying the Critical Success Factors for Information Systems to Manage Sponsored Research Activities at Institutions of Higher Education

Dwayne Lehman
lehman@rmu.edu
Information Systems and Communications

Connie Ruzich
ruzich@rmu.edu
English

Robert Morris University
Moon Township, PA 15108, USA

Abstract

Sponsored research at an institution of higher education is big business and requires an appropriate information system to meet the sponsor's fiscal and regulatory compliance standards. This research sought to identify the critical success factors for an information system to manage sponsored research at an institution of higher education and determine if there were perceived differences in the factors between department/college level and central/university level research administrators. A Delphi panel of expert research administrators with more than eight years of experience in the field of research administration who worked for highly ranked research colleges or universities (according to the Carnegie Classification) identified six critical success factors needed to manage sponsored research at an institution of higher education. The findings indicate the need for continuity of information throughout the lifecycle of a sponsored project and the integration of existing organizational information systems to manage sponsored research. These factors are important for institutions of higher education as they replace legacy systems and implement new enterprise systems to manage sponsored research. Although there were no statistically significant perceived differences of information system success factors between department/college and central university research administrators, several trends were identified. One trend identified was that department/college-level research administrators desired more financial tools to aid them in the budget development and expenditure forecasting of sponsored projects over what central/university-level research administrators indicated. A second trend identified was that department/college-level research administrators were less concerned about the technical aspects of an information system in comparison to central/university-level research administrators.

Keywords: Delphi technique, information systems, critical success factors, system integration, university research administration.

1. INTRODUCTION

Research is an integral part of the post-secondary education institution mission and represents a significant portion of all activity on college and university campuses. This study focuses on the administration and management of research activities conducted at colleges and universities. Whereas faculty members, research scientists, and other academic personnel lead these efforts

of investigation, innovation, and exploration to expand understanding of the world and develop new products and ideas, research administrators assist investigators by managing the non-scientific elements of this work. The administration of these research activities requires suitable information systems to perform the tasks related to proposal submission, contracting, personnel and financial management, and regulatory compliance. Determining the necessary elements of this information system is essential to the overall success of the research enterprise, particularly the research administrator's job of facilitating the investigator's research and the institution's adherence to applicable laws and regulations.

The purpose of this project was to determine the critical success factors necessary for an information system (IS) to effectively manage sponsored research activities at all levels within institutions of higher education. Critical Success Factors are the select areas that are essential for the success of the person, unit, or organization (Bullen & Rockart, 1981). These factors determine the health and vitality of the organization and require the manager's continual attention, support, and evaluation in order to achieve goals (Caralli, Stevens, Wilke, & Wilson, 2004; Paramenter, 2007). This project also aimed to determine if there are perceived differences between department/college level research administrators and central/university level research administrators in the necessary factors for an information system. This research adds to the body of knowledge by applying the critical success factor theory to identify the necessary factors for an information system used by research administrators within higher education to manage sponsored research. As new information systems are implemented or existing ones upgraded within an organization, the establishment of these factors will be important in designing, choosing, and evaluating these systems. The critical success factors essential for an information system to properly manage sponsored research have not been previously identified.

The administration and management of sponsored research at an institution of higher education is a multifaceted task that spans across an organization. The information system needed to serve the research administrator's needs is equally complex. A Delphi technique was used to capture and identify the critical success factors necessary for an information system to manage sponsored research across all levels of research administration within an organization. This methodology is appropriate for researching

complex issues "where large scale quantitative hard data fails to unearth the richness in tacit knowledge to help the research understand subtle expert opinion" (Grisham, 2009, p. 112). The Delphi research methodology leverages the knowledge and experiences of a select group of experts or qualified professionals to obtain a consensus on multifaceted issues through an iterative process.

This research project is limited in scope. Only institutions of higher education located in the United States (US) were examined. Other types of organizations that conduct sponsored research were excluded. This field project focused on the information systems available for the management and administration of research activities at an institution of higher education administration and does not include a discussion of electronic research administration (ERA) tools and products provided by sponsoring agencies. Although these institutions of higher education may use other information systems related to managing financial data, human resources, and student information, these systems were not evaluated. Lastly, this study focused on the end-user of the information system, the research administrator who is responsible for the administrative management of research activities within the organization. Other groups such as faculty, scientists, and other members of the organization holding administrative positions were excluded.

2. SPONSORED RESEARCH IN HIGHER EDUCATION

For the purpose of this research project, sponsored research refers to a written formal agreement entered into with external agencies that drive the financial resources of these research efforts. These agreements may appear in the form of grants, contracts, cooperative agreements, gifts, and other types of financial mechanisms (Office of Management and Budget, 2004). Each agreement that provides funding for academic research may also contain specific regulation and compliance terms and conditions. The primary agencies that provide research funding to colleges and universities include the federal government, state and local governments, private businesses, and non-profit foundations (National Science Foundation, 2015).

Sponsored research at colleges and universities is big business. According to the National Center for Science and Engineering Statistics (NCSES), universities spent over $67.3 billion on research and development in 2014, a 0.2% increase from

2013 (National Science Foundation, 2015). Research activities can bring prestige to a university or college and increase its competitive rankings and assessment (Turk-Bicakci & Brint, 2005). As a result, institutions of higher education are including research agendas as a major part of the organization's strategic plan and are seeking out new partnerships with corporations, governments, and non-profit foundations to grow their reputation (Derrick & Nickson, 2014; Turk-Bicakci & Brint, 2005). Leaders of higher education institutions are also promoting and developing more complex research strategies that include interdisciplinary, intercollegiate, and international collaborations to promote academic excellence and increase recognition and ranking (Langley & Huff Ofosu, 2007; Rutherford & Langley, 2007).

While sponsored research is vital to many higher education institutions, federal research funding is declining and is subject to tighter compliance and fiscal controls. In fiscal year 2014, the federal funding for higher education research and development dropped 5.1% after adjusting for inflation (National Science Foundation, 2015). The National Center for Science and Engineering Statistics (NCSES) reports that federal research funding to institutions of higher education fell more than 11% since 2011 and "this is the longest multiyear decline since this data started to be collected in 1972" (National Science Foundation, 2015, p. 1). Additionally, while the National Institutes of Health (NIH) reports a 40% increase in the number of research applications received since 2003, its amount of research funding has remained fairly level (Kulakowski E. C., et al. 2007). This has led to a reduction in NIH funding of all submitted research proposals from the 30th percentile range down to the 10th percentile range (Kulakowski E. C., et al., 2007). This decline in sponsored research has impacted a significant number of very high and high research institutions classified according to the Carnegie Classification of Institutions of Higher Education.

In addition, regulatory and reporting requirements by sponsoring agencies have also increased. Smith, Trapani, Decrappeo, and Kennedy (2011) state, "whereas the cost of each individual regulation may not appear to be significant, the real problem is the gradual, ever-increasing growth or stacking of regulations" (p. 57) hindering the investigator's productivity and increasing the administrative requirements of performing research. These factors have also affected the amount of time researchers can dedicate to performing research. In the 2012 Faculty Workload Survey, Schneider, Ness, Rockwell, Shaver, and Brutkiewicz (2012) report that, "researchers spend approximately 42% of their research time focused on administrative tasks such as proposal preparation, preawared and post award administration and report preparation for federally sponsored research projects instead of actually conducting research" (p. 6). Even with adequate research administration assistance, researchers stated that the administrative requirements for sponsored research projects would still consume 31% of their time (Schneider, et al., 2012).

Sponsored research is growing both in terms of the complexity of the research being conducted and in terms of the fiscal, regulatory, and contractual requirements set by sponsoring agencies. Increasing competitiveness for limited sponsored research funding adds to the complexity of managing sponsored research in higher education. The administration and management of these sponsored research activities requires a robust information system. The information systems needed to support research administration require increasingly complex project management structures and methods in order to meet demand (Rutherford & Langley, 2007). There is a critical need to better understand the essential elements of an information system that can efficiently manage the administration of sponsored research.

3. RESEARCH ADMINISTRATION

The administration of research conducted at institutions of higher education represents the business support necessary for the success of any exploratory initiative (Kulakowski & Chronister, 2006). The increased competition for limited research funding, combined with the sponsors' demand for tighter fiscal accountability and reporting requirements, has expanded the role and responsibilities of the research administrator (Lintz, 2008). Today research administrators (RA) are fully integrated throughout all levels of the organization, perform a diverse collection of duties and require a working knowledge of the legal, ethical, scientific, and fiscal components of academic research (Lintz, 2008; Shambrook & Roberts, 2011). Figure 1 (Appendix A) illustrates the various roles and responsibilities of research administrators.

The administration of a sponsored project can be categorized into two primary areas: pre-award and post-award (Kulakowski & Chronister, 2006; NCURA, 2015). The pre-award project activities include finding funding, proposal development,

proposal submission, award negotiation, and project setup (Campbell, 2010; Kulakowski & Chronister, 2006; NCURA, 2015). The post-award administration activities include accounting, accounts payable and receivable, property and inventory control, payroll and reporting (Campbell, 2010; Kulakowski & Chronister, 2006; NCURA, 2015). Dependent on the amount of research conducted at an institution of higher education and the organizational structure of personnel and authority, and research administration function, research administration can be further divided into two general groups: research administration and management at the university or central level, and the administration and management of research that operates at the department/college level (Campbell, 2010). The central-level research administrators primarily have an external focus and a broadly defined responsibility to ensure that the institution promotes excellence in the conduct of research (Galland, McCutcheon, & Chronister, 2008). These research administrators represent the organization and are primarily concerned with compliance (NCURA, 2015). The department/college-level research administrators are typically more internally focused and are primarily concerned with the direct support of the researcher, the responsibilities of others working on the project, and the academic department to which the researcher is assigned (Campbell, 2010; Kulakowski & Chronister, 2006; NCURA, 2015). The identification of critical success factors for an appropriate information system is essential to the overall success of the administrator's job performance, the facilitation of the investigator's research, and the regulatory management of the institution.

4. CRITICAL SUCCESS FACTOR THEORY

The Critical Success Factor Theory was first introduced and applied to the field of information systems by John Rockart (1979) in the *Harvard Business Review* (HBR) article "Chief Executives Define their own Data Needs" based on D. Ronald Daniel's 1961 article "Success Factors" (Rockart, 1979 p. 85). Rockart defined critical success factors as:

> The limited number of areas in which satisfactory results will ensure successful competitive performance for the individual, department, or organization. Critical success factors are the few key areas where 'things must go right' for the business to flourish and for the manager's goals to be attained (Bullen & Rockart, 1981, p.7).

Rockart initially developed the critical success factor (CSF) theory as a management information system (MIS) planning tool to identify and communicate a manager's information requirements (Boynton & Zmud, 1984). These management-identified factors or elements are vital to the organization's evolution and must receive constant attention, support, and evaluation. The primary advantage of the critical success factor (CSF) theory is that it communicates and makes explicit the major concerns of management, thus reducing organizational ambiguity (Boynton & Zmud, 1984; Caralli et al., 2004). Paramenter (2007) states, "Better practices suggests that there should be only between five and eight CSFs" (p. 29). The critical success factor theory is an established information systems tool to discover and communicate the information system requirements at various managerial levels. This study focuses on applying the critical success factor theory to a profession (research administrators) within the industry of higher education.

5. METHODOLOGY

An exploratory sequential mixed-methods research strategy was used for this project. The data collection method followed a modified e-Delphi technique for ranking-type surveys and consisted of two phases (Keeney, et al., 2011). Both phases were conducted through an on-line survey. The initial, qualitative phase sought to discover the issues, and aimed to gather consensus around the most important issues (Keeney, et al., 2011; Schmidt, 1997). The subsequent stages sought to prioritize or rank these issues (Abu, et al., 2012; Schmidt, 1997).

The Delphi research methodology leverages the knowledge and experiences of a select group of experts or qualified professionals to obtain a consensus on multifaceted issues through an iterative process. This methodology is appropriate for researching complex issues "where large scale quantitative hard data fails to unearth the richness in tacit knowledge to help the research understand subtle expert opinion" (Grisham, 2009, p. 112). There are four goals associated with a Delphi study: (1) gather and summarize knowledge from an expert panel, (2) obtain an agreement or consensus in regard to the topic or issue, (3) explore ideas with knowledgeable participants, and (4) provide information to aid in decision-making (Abu, Ritchie, & Jones, 2012). The Delphi Method is systematic, flexible, and allows for the use of a variety of communication methods and tools

(Abu, et al., 2012; Dalkey, 1969). It is also insightful and produces reliable and valid results (Abu, et al., 2012; Dalkey, 1969; Grisham, 2009).

A purposive sampling technique was employed for this study. A study utilizing the Delphi technique can have any number of participants. The ideal sample size of experts is large enough to represent the population, conduct the desired research, and yet remain manageable by the researcher (O'Leary, 2014; Williams, 2004). Okoli and Pawlowski (2004) recommend an expert panel size of between 10 and 18 participants (p. 19). The participants for this study consisted of research administrators (RA) from Very High and High Research Institutions, as identified by the Carnegie Classification of Institutions of Higher Education, and participants were required to have a minimum of eight years experience in the field of research administration; they had to be familiar with both pre-award and post-award research activities, and they were required to be currently using an institutional information system to manage sponsored research.

The Delphi panel consisted of 18 people. Fifteen females and three males completed both surveys (Figure 1). The gender percentages for this Delphi study are consistent with the 2010 Profile of a Research Administrator (Shambrook & Roberts, 2011). Sixty-seven percent of the participants had 17 or more years in the field of research administration and 33% had between 8 and 16 years of RA experience.

RAs working at the central/university level represented 78% and department/college level RA represented 22% of the Delphi panel (Figure 2). According to the 2010 Profile of a Research Administrator, 30.3% of research administrators identified themselves as working at the department level (Shambrook & Roberts, 2011).

One limitation of this research is that the number of participants identifying themselves as department/college level research administrators did not match the 2010 RA profession profile; over representing research administrators at the central/university level. Seventy-two percent of the participants worked at Very High research institutions, while 28% indicated they worked at an institution with High research activity according to the Carnegie Classification of Institutions of Higher Education. Tables 1 and 2 indicate the state and institutional control of the institutions the Delphi participants identified as working for.
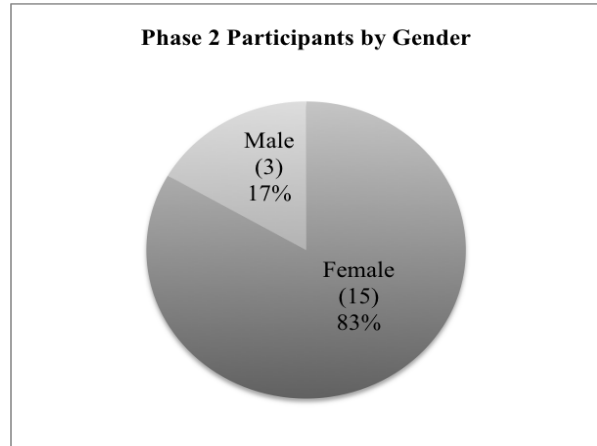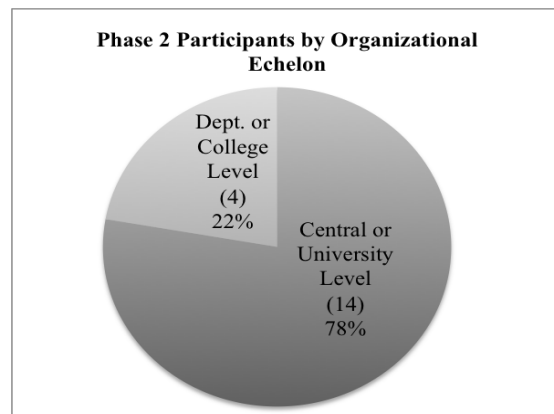

Figure 1: Phase 2 participant gender

Figure 2: Phase 2 participant organizational

Figure 3: Phase 2 participants institutional research activity

Table 2: Carnegie Classification High research institutions from the following states were represented

High Research Institutions	
State	**Control**
Alaska	Public
California	Public
Illinois	Private not-for-profit
Maryland	Public
Mississippi	Public
North Carolina	Public
Ohio	Public

Table 3: Carnegie Classification Very High research institutions from the following states were represented

Very High Research Institutions	
State	**Control**
California	Private not-for-profit
California	Public
Florida	Public
Kentucky	Public
Maryland	Private not-for-profit
Massachusetts	Private not-for-profit
Montana	Public
Oklahoma	Public
Pennsylvania	Private not-for-profit
Pennsylvania	Public

6. RESULTS AND DISCUSSION

Phase 1 survey results
Upon completing the analysis of the Phase 1 survey results, the following 22 factors (Table 4, Appendix A), listed in the order of frequency, were identified as being critically important for an information system to manage sponsored research at an institution of higher education. Additionally, the most common theme that emerged was the issue concerning the efficacy and number of information systems used to manage sponsored research. Eighty percent of the department/college level research administrators and 59% of central/university level research administrators indicated that separate, nonintegrated, and inadequate information systems were the most frequent problem or obstacle in the management of sponsored research.

The use of proprietary or homegrown information systems was prevalent among the Delphi panel. Fifty percent indicated the use of an organization-created information system(s) to meet their needs. Twenty-seven percent stated that they used a shadow system in addition to the information systems provided by the institution. For the purposes of this research, a shadow system is defined as any spreadsheet application or database that replicates data and functionality of an organizational information system to address the deficiencies of the existing information system (Behrens, 2009). Lastly, the participants indicated that the problems with research administration information systems spanned the lifecycle of sponsored research from pre-proposal development, through award set-up and post-award management, to reporting and project closeout.

Phase 2 survey results
Six of the 22 factors identified by the Delphi panel participants for an information system to manage sponsored research activities at an achieved greater than a 50% majority, had a mean greater than 4.40, had a standard deviation of less than 1.00, and attained an IQR of 1.00 or less. One of the goals of the Delphi methodology is to obtain consensus among the participants, these qualifying measurements were chosen because they indicate a high level on consensus within the Delphi panel. Those factors were
1. Must be accessible through the Internet/Intranet
2. Must have top leadership support
3. Must be easy to use (user friendly) for different users at all levels
4. Must work and integrate across existing institutional information systems and platforms
5. Must have dedicated, continual IT support
6. Must be able to attach, store, and retrieve supporting documentation

Two of the six critical success factors needed for an information system to manage sponsored research (must work and integrate across existing institutional information systems and platforms and must be able to attach, store, and retrieve supporting documentation) were related to the data management functionality of an information system. The first CSF in this category was the ability to work and integrate across existing institutional information systems and platforms. Dowdy and Schultz (2015) provide a

possible explanation for the lack of system integration by stating,

> Generally speaking, yesterday's legacy systems were often stovepipe applications. That is, the pre-award system met the needs of the pre-award office but had little or no interaction with the financial system. The financial system had little to no interaction with the human subjects system or the intellectual property system. In effect, each ERA process was a stand-alone application, developed over time, to satisfy a particular business processes or transaction (p. 905-6).

Data management functionality is also key in the sixth ranked CSF: the ability to be able to attach, store, and retrieve supporting documentation. The amount of documentation and records generated from pre-award proposal development through contract negotiations and award set-up to project closeout is substantial. A single document repository might logically increase efficiency and improve communication. Sponsored research projects can have durations ranging from a few months to a decade or longer. The importance of capturing the continual flow of documentation and information for a sponsored project throughout its lifecycle in a single information repository could prevent the loss of information through personnel transition and turnover, and aid in the accuracy, timeliness, and completeness of project reporting.

The 4th ranked CSF, that of being easy to use (user friendly) for different users at all levels, and the 1st CSF, that of being accessible through the Internet/Intranet, were related to the user interface of the information system. Ease of use focuses on the end-user requirement for the information system, which is consistent with the rules of user-centered design (Norman, 1988). Finally, organizational factors were key to two CSF's and ranked 2nd and 5th. The participants considered reliability and consistent operational access to the sponsored projects information system the top priority through the existence of continual, dedicated IT support. This is consistent with the participants' concern for information system reliability and demonstrates their dependence on information systems to perform the duties associated with research administration. The other organizational factor considered critical was the support from top leadership. Top leadership support can be interpreted as support from the President, Provost, or Vice-Provost of the institution to provide funding and resources in the investment, maintenance, and upgrade of research administration information systems. However, it can also be interpreted as the advocacy, moral encouragement, and championing of the research administrator and the important contributions of his/her work to the research mission. Further research is needed to clarify the role top leadership plays in sponsored research administration.

The cross-tab analysis of Likert-type scale responses to the 22 factors did not yield any statistically significant findings between the central/university-level and the department/college-level of research administrators. When the comparing the ranking of extremely important factors between the two groups were compared, no statistically significant findings were found; however, several trends were discovered. On average, department/college level participants ranked three specific factors in the top 5 more frequently than central/university level participants on the Likert-type scale. The factors include the following:

- Must provide budget forecasting tools
- Must provide budget-development tools
- Must provide automated effort certification reporting tools

One hundred percent of the department/college level participants indicated that budget forecasting tools were *extremely important* or *very important* for an information system to manage sponsored research. However, only 35.7% of the central/university level participants indicated this factor as either extremely or very important. Likewise, 75% of the department/college level participants indicated that a budget development tool was extremely important for information systems managing sponsored research. Only 14.2% of central/university level participants indicated this factor was extremely important. One possible explanation for this trend is that department/college level research administrators work more closely with faculty researchers and often prepares the project budget for the proposal submission based on the researcher's guidance. Department/college level participants had a tendency to be less concerned about the technical aspects of information systems used to manage sponsored research. Again, there were no statistically significant findings. When the ranking of extremely important factors was compared, the following five technical factors for an information system to manage sponsored research were observed to be consistently ranked as less important for department/college level participants than for central/university level participants:

- Must be able to expand capabilities based on institution needs
- Must be easy to update based on policy and regulation changes
- Must provide continuity of information from pre-award proposal development through post award management
- Must be accessible through the Internet / Intranet
- Must have dedicated, continual IT support

One possible explanation for this observation is that typically the Vice President of Research, or its equivalent title, is a central/university level position and it is this position that is responsible for the research administration systems and technology used at an institution (NCURA, 2015). Department/college level research administrators may accept a fatalistic perspective with regard to the technical capabilities of an institution's information system to manage sponsored research.

The need for ERP ERA
The data suggests that an enterprise-level information system solution to manage sponsored research at an institution of higher education could be advantageous. Two of the critical success factors are related to the integration and communication of information across an institution of higher education (must work and integrate across existing institutional information systems and platforms and must be able to attach, store, and retrieve supporting documentation). Additionally, three factors were determined by the Delphi panel to be *extremely important* and achieved a greater than 50% majority, but not considered critical to the management of sponsored projects also address the need for system integration. The three factors that promote the adoption of an integrated ERP are as follows:

1. Must provide continuity of information from pre-award proposal development through post award management
2. Must be able to provide data analytics for robust and flexible reports at all levels across the organization
3. Must be able to monitor and track compliance requirements (IRB, IACUC, COI, etc.)

From an administrative perspective, the need for system integration for a research administration information system seams clear: 86.4% of the participants indicated that they use one to nine separate systems, and on average, participants reporting using 3.13 separate information systems to perform their duties as research administrators. Additionally, 27% of the participants indicated that they use a shadow system in addition to organization-provided systems to perform their jobs. Implementing an enterprise resource planning (ERP) for the administration of sponsored research has the possibility of decreasing the redundancy of data entry, as well as improving administrator efficiency and accuracy.

From a business perspective, the proper and effective management of sponsored research is critical to the financial sustainability and success of the institution. For example, at Carnegie Mellon University (CMU) sponsored research makes up 33.9% of the annual operating revenue (Carnegie Mellon University, 2015). Likewise, the University of California (UC), a large, public state university system with numerous campuses that includes UC Berkley, UC Davis, UC Irvine, UC Los Angeles, and others, received approximately $5.44 billion in sponsored research funding in fiscal year 2014, which represents slightly over 20.4% of their annual revenue (University of California, 2015). Given the significant percentage of operating revenue generated from sponsored research for the Carnegie Classification Very High and High research institutions, an ERP-level information system is recommended.

Challenges of ERP implementation
The implementation of an ERP research administration system is a challenge for institutions of higher education. The major challenges for these institutions are providing the dedicated financial infrastructure and human resources necessary to accomplish the ERP implementation. Electronic Research Administration (ERA) systems require a substantial initial capital investment from the organization and continued monetary support for the maintenance and upkeep of the ERP system. Additionally, the ERP system will require significant time to phase out existing legacy systems, train staff and faculty on the new system, and design and adopt new business processes.

The organizational culture is also a major challenge with implementing an ERP system at institutions of higher education. Colleges and universities are predominately structured in silos (Chisita & Abdullahi, 2012; Evans & Malina, 2010; Kolowich, 2010). For the purposes of this study, a silo is defined as an organizational structure that promotes departmentalization and specialization within different units of the organization (Chisita & Abdullahi, 2012; Evans & Malina, 2010; Kolowich, 2010). One reason for this phenomenon is that it promotes a "strong college model . . . which emphasizes the

individual brands of different colleges on a campus [and] empowers those schools to attract talented scholars and funding for important research in their particular disciplines" (Kolowich, 2010, p. 1). However, this structural model also creates communication, interdisciplinary research, and information technology challenges (Kolowich, 2010).

Another cultural challenge is the actual value and importance leadership places on research administration within the organizations. The use of multiple proprietary and legacy systems used to manage sponsored research suggests a possible disconnect between the espoused value concerning the importance of research administration and the actual actions leaders take to support and champion the administrative tasks associated with sponsored research. The lack of dedicated resources to support the implement an ERA system could indicate a divide between espoused values and lived values. The benefits of implementing an enterprise-level ERA information system have the ability to lessen the obstacles associated with the archetypal departmental structure of colleges and universities and promote more effective communication, knowledge sharing, and cross-disciplinary institutions.

7. CONCLUSION

Sponsored research at an institution of higher education is big business. This research focuses on applying the critical success factor theory within the industry of higher education to identify the critical success factors for the information systems used by research administrators professionals at all levels within the organization to successfully manage sponsored research. This study identified six critical success factors for an information system to manage sponsored research at an institution of higher education. Although no clear statistically significant findings were evident between department/college level and university/central level research, administrator information system needs were identified and several trends were observed. The list of identified critical success factors should not be considered exhaustive, but rather viewed as a door to other research opportunities in the research administration profession and higher education information systems.

8. REFERENCES

Abu, E. E., Ritchie, C., & Jones, E. (2012). Consulting the oracle? *International Journal of Contemporary Hospitality Management, 24* (6), 886-906.

Behrens, S. (2009). Shadow systems: the good, the bad and the ugly. *Communications of the ACM, 52* (2), 124-129.

Boynton, C. A., & Zmud, W. R. (1984). An assessment of critical success factors. *Sloan Management Review*, 17-27.

Bullen, C. V., & Rockart, J. F. (1981). *A primer on critical success factors.* Massachusetts Institute of Technology, Sloan School of Management. Cambridge: Center for Information Systems Research.

Campbell, D. R. (2010). *The role and development of the research administration profession in higher education.* Washington State University, Department of Educational Leadership and Counseling Psychology. Seattle: Washington State University.

Caralli, R., Stevens, J., Wilke, B. J., & Wilson, W. R. (2004). The critical success factor method: a foundation for enterprise security management. *CMU/SEI-2004-9TR-010* .

Carnegie Mellon University. (2015). *2014-2015 Annual Report Carnegie Mellon University*. Retrieved from https://www.cmu.edu/finance/reporting-and-incoming-funds/financial-reporting/files/2015_annual-report.pdf

Chisita, C. T., & Abdullahi, I. (2012). Rising above the grain silo mentality through collaboration: creating opportunities between the LIS educators and practioners in developing countries. *World Library and Information Conference* (pp. 1-16). International Federation of Library Associates.

Dalkey, N. (1969). *The Delphi method: an experimental study of group opinion.* The Rand Corporation. Santa Monica: Rand.

Derrick, G., & Nickson, A. (2014). Invisible intermediaries: a systematic review into the role of research management in university and institutional research process. *Journal of Research Administrators, XLV* (2), 11-45.

Dowdy, S. D., & Schultz, L. (2015). Electronic research administration. *Sponsored Research Administration: a guide to effective strategies and recommended practices*, 905-1 - 905-19.

Evans, K., & Malina, R. (2010). Breaking down the silos: curriculum development as a tool for crossing disciplines in the arts, sciences, and humanities. *Educause Mid-Atlantic Regional Conference* (pp. 1-8). Inside Higher Ed .

Galland, J. C., McCutcheon, J. R., & Chronister, L. U. (2008). Laboratory Management Institute: A model for the professional development of scientists. *The Journal of Research Administrators, XXXIX* (2), 51-67.

Grimshaw, S., & Wilson, I. (2009). Establishing user needs: a large-scale study into the requirements of those involved in the research process. *The Journal of Research Administration, XL* (1), 32-48.

Grisham, T. (2009). The Delphi technique: a method for testing complex and multifaceted topics. *International Journal of Managing Projects in Business, 2* (1), 112-130.

Holland, P. C., & Light, B. (1999). A critical success factors model for ERP implementation. *IEEE*, 30-36.

Kamaruddin, M., Razali, R., & Deraman, A. (2011). Critical success factors of executive information systems development for educational management: a preliminary investigation. *2011 International Conference on Electrical Engineering and Infomatics* (pp. 1-6). Bandung: IEEE.

Keeney, S., Hasson, F., & McKenna, H. (2011). *The Delphi technique in nursing and health research.* West Sussex, UK: Wiley-Blackwell.

Kolowich, S. (2010, January). Blasting academic silos. *Inside Higher Ed* .

Koumaditis, K., Themistocleous, M., & Da Cunha, P. R. (2013). SOA implementation critical success factors in healthcare. *Journal of Enterprise Information Management, 26* (4), 343-362.

Kulakowski, E. C., & Chronister, L. U. (2006). *Research administration and management.* Sudbury, MA: Jones and Bartlett Publishers.

Kulakowski, E. C., Chronister, L., Molfese, V., Slocum, J. M., Studman, C., & Waugaman, P. (2007). Voice of experience. *The Journal of Research Administration, XXXVIII*.

Langley, D., & Huff Ofosu, M. (2007). Celebrating a profession: the global perspective. *Journal of Research Administration, XXXVIII*.

Linstone, H., & Turoff, M. (2002). *The Delphi method: techniques and applications.* Retrieved from http://is.njit.edu/pubs/delphibook/delphibook.pdf

Lintz, E. M. (2008). A conceptual framework for the future of successful research administrator. *The Journal of Research Administration, XXXIX* (2), 68-80.

National Institutes of Health. (2015). *U.S. Department of Health & Human Services*. Retrieved from http://report.nih.gov/funded_organizations/index.aspx

National Science Foundation. (2015). *Universities Report Continuing Decline in Federal R&D Funding in FY 2014.* Retrieved from http://www.nsf.gov/statistics/2016/nsf16302/

NCURA. (2015). *Level 1: fundamentals of sponsored project administration workshop 2.0.* Washington, DC: National Council of University Research Administrators.

Norman, D. A. (1988). *The design of everyday things.* New York, NY: Basic Books.

Office of Budget and Management. (2015). *2 CFR 200 Uniform Administrative Requirements, Cost Principles, and Audit Requirements for Federal Awards.* Retrieved from https://www.gpo.gov/fdsys/granule/CFR-2014-title2-vol1/CFR-2014-title2-vol1-part200/content-detail.html

Office of Management and Budget. (2004). *Circular A-21*. Retrieved from http://www.whitehouse.gov/omb/circulars/a012/a21_2004.html

Okoli, C., & Pawlowski, S. D. (2004). The Delphi methods a research tool: an example, design considerations and application. *Information and Management, 42* (1), 15-29.

Paramenter, D. (2007). *Key performance indicators: developing, implementing and using winning KPIs.* Hoboken, NJ: John Wiley and Sons.

Program Management Office. (2015). *About the Grants.gov program management office*. Retrieved from http://www.grants.gov/web/grants/about.html

Rockart, J. F. (1979). Chief executives define their own data needs. *Harvard Business Review* , 81-93.

Rutherford, S., & Langley, D. (2007). Implementation of systems to support the management of research: commentary from a U.K. university perspective . *The Journal of Research Administration, XXXVIII*, 85-95.

Schmidt, R. C. (1997). Managing Delphi surveys using nonparametric statistical techniques . *Decision Sciences, 28* (3), 763-774.

Schmidt, R. C., Lyytinen, K., Keil, M., & Cule, P. (2001). Identifying software project risks: an international Delphi study. *Journal of Management Information Systems, 17* (4), 5-36.

Schneider, S. L., Ness, K. K., Rockwell, S., Shaver, K., & Brutkiewicz, R. (2012). *2012: Faculty workload survey research report.* Federal Demonstration Partnership (FDP).

Shambrook, J., & Roberts, T. J. (2011). 2010 Profie of a Research Administrator. *Research Management Review, 18* (1), 19-30.

Smith, T. L., Trapani, J., Decrappeo, A., & Kennedy, D. (2011). Reforming regulation of research universities. *Issues in Science and Technology, 27* (4), 57-64.

Turk-Bicakci, L., & Brint, S. (2005). University-industry collaboration: patterns of growth for low-and middle-level performers level. *Higher Education, 49* (1-2), 61-89.

U.S. Department of Education. (2014). *National Center for Educational Statistics*. Retrieved from https://nces.ed.gov/programs/digest/d14/tables/dt14_317.10.asp?current=yes

University of California System. (2015). *University of California Financial Reports.* Retrieved from http://finreports.universityofcalifornia.edu/index.php?file=retrends/retrends_2014.pdf

Appendix A

Strategic	**Policy**	**Operational**
Research strategy	Research Ethics	Proposal development
Research theme development	Research governance and integrity	Budgeting
Portfolio management	Reporting	Proposal submission
International research	Intellectual property	Contract negotiation and monitoring
Trend analysis	Technology transfer	Post award financial management
Business development	Start-ups and commercialization	Project Management
Risk assessment and monitoring	Auditing	Clinical trials and research
(Adapted from Langley & Huff Ofosu, 2007; Lintz, 2008; Kulakowski & Chronister, 2006)		

Table 3: Duties and Responsibilities of research administrators

Planning and Preparation

- Define research problem
- Develop research questions and design
- Develop criteria for participation

Phase 1: Qualitative Data collection

- Expert panel selection
- Discovery of Issues
- Analyze responses

Phase 2: Quantitative Data Collection

- Prioritize findings from previous phase
- Rank the most important factors
- Analyze responses

Results

- Consensus of expert panel
- Reporting of findings

Figure 4: Research design

1	Must be able to provide data analytics for robust and flexible reports at all levels across the organization
2	Must be easy to use (user friendly) for different users at all levels
3	Must provide automated error checking of proposals to ensure compliance
4	Must work and integrate across existing institutional information systems and platforms
5	Must have dedicated, continual IT support (either in-house or provided by vendor)
6	Must be able to expand capabilities based on institution needs
7	Must provide continuity of information from pre-award proposal development through post award management
8	Must be capable of electronic proposal submission to prime or sponsoring agency (system-to-system capability)
9	Must be able to monitor and track compliance requirements (IRB, IACUC, COI, etc.)
10	Must provide electronic processing of proposal applications (Internal Routing and approvals)
11	Must be easy to update based on policy and regulation changes
12	Must be accessible through the Internet / Intranet (web-based or web enabled)
13	Must be able to link to sponsor or prime agency regulation and guideline references
14	Must have an alert mechanism to identify upcoming proposal deadlines, budget and expense variances, and compliance requirements
15	Must be able to attach, store, and retrieve supporting documentation
16	Must provide budget development tools
17	Must provide real-time expenditure tracking
18	Must provide budget forecasting tools
19	Must be able to provide customized dashboards for all users
20	Must provide automated effort certification reporting tools
21	Must be able to monitor and track sub-awards
22	Must have top leadership support

Table 4: 22 Factors for an information system to manage sponsored research at an institution of higher education

Building Client Vendor Alignment Capability in Strategic Information Systems Outsourcing

Biswadip Ghosh
bghosh@msudenver.edu
Computer Information Systems
Metropolitan State University of Denver
Denver, Colorado 80217, USA

Abstract

Strategic information systems outsourcing (ISO) refers to the outsourcing of information systems that are anticipated to have a major transformational impact on the client's business strategy. Such outsourcing arrangements typically span longer terms and pose higher risks for failure. These risks need to be mitigated by building inter-firm alignment, which help define roles initially and sustain responsibilities over the long term. Inter-firm alignment capability keeps the client vendor relationship going thru knowledge sharing in strategic planning and business and information technology operational processes. This research study defines a multi-item measure of client-vendor alignment capability and uses that instrument to survey a number of North America based firms, who have undertaken the outsourcing of strategic information systems to their Indian vendor and finds that both contractual and relational governance are needed to build this inter-firm alignment capability. The results indicate that both contractual and relational governance support knowledge sharing, which builds client-vendor alignment and this alignment capability impacts strategic outsourcing success factors.

Keywords: Information Systems Outsourcing, Alignment, Contractual and Relational Governance, Knowledge Sharing, Success Factors

1. INTRODUCTION

Information Systems Outsourcing (ISO) refers to transferring the development, provisioning and/or support of IS/IT products or services to a vendor for an agreed upon time, cost and functional scope (Dibbern, et.al. 2004). Due to the recent movement towards utilizing hosted information systems (IS) and cloud-based system providers, worldwide outsourcing spending has grown to over $800 Billion a year. Traditionally past IS outsourcing has focused on IT staff cost reduction and divesting non-core, secondary value-chain activities of the client such as payroll or help-desk systems. However, the current and growing trend in ISO is the pursuit of strategic deals that are intended to be transformational for the client's business. Typical objectives of strategic outsourcing transcend cost savings and include adopting novel systems that can shift the client organization's competitive position, enhance core competencies of the client, creating value, increasing flexibility to meet changes in future business conditions and exploiting new markets (Grant, 2003; Greaver, 1999). Outsourcing deals can also pose severe risks such as loss of control, have hidden costs, business uncertainty and lead to erosion of client knowledge and have the potential for systems failure (Earl, 1996). Recent examples of strategic ISO include the 10 year 20M pound deal between Northern Ireland's social security agency and EDS and PricewaterhouseCoopers, the 7 year 28M pound deal between Ofstead and Logica for the Early Years Education initiative (Phillips, 2014) and Lufthansa's 7 year 70M Euro deal with IBM Global Services to outsource its IT infrastructure services (Flinders, 2014).

Characteristics of Strategic ISO
This growth trend of strategic ISO is currently being seen in the North American energy exploration industry, where medium sized exploration companies are pursing business transformation thru the sourcing of strategic information systems using long term deals with offshore IS vendors. An example is the sourcing of an enterprise resource planning system to

manage the mining field sites and optimizing exploration activities for an oil and gas company. This outsourcing is highly strategic as it impacts the client's business for many years into the future. The benefits to clients come from utilizing and leveraging the knowledge of external vendors (Chang and Gurbaxini, 2012) in their IS projects, adopting the latest IS project methodologies, improving internal business processes, getting access to trained and experienced IS staff from the vendor and eliminating the overhead of having to frequently upgrade in-house technology infrastructure and system components. As outsourcing moves to this next level, clients seek greater value and diverse objectives (Mukherjee, et.al, 2013). However, the realization of broader benefits is contingent upon the client and vendor firms' ability to synergistically manage their resources and build inter-firm capabilities in a dynamic environment. This requires sophisticated vendor management activities that rely on elements of both contractual and relational governance (Willcocks, et.al.,1999; Rottman&Lacity, 2004). Strategic ISO deals are less formulated at first and require multiple planning cycles, frequent readjustment of priorities and redefinition of architecture and roles. More advanced client-vendor inter-firm capabilities to help manage complex boundary spanning systems development processes and fostering collaboration are needed to co-create substantial value in the ISO relationship over time (Rai, et.al., 2012).

Client Vendor Alignment in Strategic ISO

The success of such strategic outsourcing deals depends on the sharing and transformation of knowledge over the long term between the client and vendor. To achieve this, client vendor alignment needs to be established which aligns each firm's objectives, resources and processes and builds consensus on the opportunities and challenges facing the deal (Klein and Rai, 2009). At the strategic level, this alignment involves linking strategic intent through the joint process of identifying core and non-core business areas. At the tactical level, the client and vendor must facilitate knowledge exchange about their management methods and values and jointly define their business processes and organizational structures for the operational aspects of the deal. Key decision makers in both organizations must be identified along with intersecting procedures in IT processes to effectively manage these strategic projects over the long term. Without adequate knowledge sharing about the strategic intent and efforts to align IT processes to connect people to people, the client and vendor can get out of sync over the course of the deal as circumstances change, causing significant sourcing issues (Rottman and Lacity, 2004; Lacity and Willcocks, 1998).

Despite these overwhelming arguments for the need to establish client vendor alignment in strategic ISO, current IS research has not addressed these questions of defining the components of inter-firm alignment or the means of using outsourcing governance mechanisms to create client vendor alignment (CVA) and the impacts that CVA may have on outsourcing success factors.

Research Goals

Business IT alignment has been recognized for several years as an important organizational capability (Luftman and Brier, 1999), but has not been studied in the context of strategic ISO. A client-vendor alignment (CVA) capability over three dimensions: strategic ("planning"), structural ("execution") and relational ("norms") can impact outcome success factors in strategic ISO scenarios. This research studies strategic ISO between an Indian vendor and several medium sized North American firms in the oil and gas and energy exploration industry. The goals of this research study are to:

1. Build a measurement model for client vendor inter-firm CVA capability.
2. Determine if CVA capability impacts the success factors of strategic ISO.
3. Determine the contributions of both contractual and relational governance on client-vendor knowledge sharing and the CVA capability.

2. THEORETICAL BACKGROUND

Strategic information systems outsourcing implies a long term commitment between the client and vendor to define, build and deploy needed solution components in an iterative manner to support business strategy goals (Greaver, 1999). Published research reports that poor vendor management practices can amplify outsourcing risks over the longer term and lead to poor performance (Lacity and Hirschheim, 1993). In a strategic outsourcing agreement, flexibility and adaptation are important to deal with future uncertainties (Kern and Willcocks, 2000). Using a multi theoretic approach, Gottschalk and Solli-Saether (2005) identified eleven critical success factors for IT outsourcing including definition and needs management, resource exploitation across the alliance, cost reduction, relationship exploitation, vendor behavior control and stakeholder management.

Types of Outsourcing Governance

Outsourcing relationships are defined and managed through the establishment of two forms of governance structures: (a) contractually defined formal controls or service level agreements (SLA) (Goo, 2010) and (b) relational mechanisms that emphasize cooperation over the long term (Kishore, et.al., 2003). Formal controls driven by written contracts help define roles and demarcation of process responsibilities across the client and vendor (Goo, 2010). They guide vendor behavior towards desired objectives if those objectives are easy to understand and fall early on in the deal (Goo, et.al., 2009). Relational governance refers to establishing a set of norms for cooperation among client and vendor that that facilitate accepting, sharing and delivering on responsibilities from the outsourcing deal. Sophisticated, long term arrangements like strategic outsourcing requires both approaches to mesh in a hybrid fashion. Studies have shown that both forms of governance are related as definition of roles from formal agreements can set the tone for relational commitment and communication channels, which are utilized to build and maintain inter-firm capabilities such as knowledge sharing (Poppo and Zenger, 2002; Goo and Huang, 2008).

A variety of inter-firm capabilities have been reportedly used to improve outsourcing outcomes (Kern &Willcocks., 2000; Palvia, et.al., 2010; Plugge, et.al., 2013). Some of the pertinent inter-firm capabilities are based on the appropriate definition of roles and process ownerships among the client and vendor and can include management ability, processes for needs definition and prioritization and integrated client-vendor systems lifecycle processes. Contracts along with IT resources that allow for creation, storage and sharing of knowledge help build these capabilities (Wade and Hulland, 2004). Inter-firm capabilities provide clear standards of operation so that when conflicts arise, the teams can work through them (Goo and Huang, 2008).

Alignment Capability

Business –IT alignment refers to the capability to apply IT in an appropriate and timely way and in harmony with business strategies (Luftman and Brier, 1999). Prior research has identified three dimensions of Business/IT alignment: (1) strategic alignment, (2) structural alignment and (3) relational alignment. Strategic alignment provides the fit between the priorities and activities of the vendor IS function and those of the client business units, so that IS and applications can be aligned with business needs. Structural alignment defines the formal organizational structures that enable the alignment of the planning, decision-making, reporting and other project management aspects between client and vendor. Relational alignment refers to the informal organizational structures, norms and agreed processes, divisions of work, formal and informal teamwork, and working relationships between the firms. Relational alignment lays the foundation for strategic and structural alignment (Ghosh and Scott, 2009).

3. DEFINITION OF STUDY CONSTRUCTS

There are five research constructs in this study. Relational governance and contractual governance are the two independent variables. They impact client vendor knowledge sharing and client vendor alignment capability. The dependent variable for the study is ISO success factors.

Contractual and Relational Governance

Contractual governance of ISO refers to establishing provisions for controlling vendor actions and is based on control theory (Goo, 2010). Outsourcing contracts establish service level agreements (SLA) that define detailed actions that the two parties will engage in during the term of the outsourcing. Important themes in outsourcing contracts include methodology, process ownership, change management, performance measurements and rewards and penalties. While contracts attempt to get as detailed and specific as possible, unforeseen circumstances can arise that may be beyond the contract scope and require other relational governance. Relational governance refers to establishing a set of norms for cooperation among client and vendor that can help the resolution of unforeseen issues during the outsourcing term. Relational governance stresses the importance of client vendor trust to foster exchange of opinions and an environment for collaborations and co-creation (Goo and Huang, 2008).

Knowledge Sharing Capability

In outsourced IS development projects, there are three parties involved that need to share knowledge – the business users in the client and the two IT organizations - one in the client and the other in the vendor. If sufficient interactions and knowledge sharing is not fostered among these three project stakeholders then poor project outcomes have been reported particularly in complex projects (Carlile, 2004). This knowledge can be related to either the information systems being sourced or the processes by which the system is being defined and developed. Drawing on the research stream

in knowledge management capabilities, client vendor knowledge sharing is defined as the availability of channels and human protocols to share project knowledge (Tanriverdi, 2005). The actual mechanism of knowledge sharing consist of one or both parties seeking knowledge and/or providing knowledge in response to a request for knowledge, such that the work of one or both parties are affected by the shared knowledge. The facets of knowledge sharing in a ISO are: (1) one or both parties seeking to acquire knowledge, (2) one or both parties converting tacit knowledge or pointing to the location of already explicit knowledge in response to the request, (3) one or both parties transferring the knowledge synchronously or asynchronously and (4) the seeking party applying the new knowledge (Ko, Kirsch and King, 2005).

Client Vendor Alignment Capability
Outsourcing of Information systems development is a knowledge intensive activity that demands coordination, communications and alignment between the client and vendor. This alignment is defined in 3 levels – at the strategic level, where prioritization decisions are made, at the operational level where these decisions are realized and at the relational level where staff have norms to work together. The client vendor alignment capability is defined as the combination of (a) strategic decision making alignment on ISP project priorities and (b) the operational process connections with systems definition, development, deployment and support (Chan, 2002). Strategic alignment provides the fit between the priorities and activities of the vendor IS function and those of the client business units, so that IS and applications can be aligned with business needs. Structural alignment defines the formal organizational structures that enable the alignment of the planning, decision-making, reporting and other project management aspects between client and vendor. The third component of alignment capability is the relational aspects, where the cross organizational teams develop norms of teamwork. Relational alignment refers to the informal organizational structures, norms and agreed processes, divisions of work, formal and informal teamwork, and working relationships between the firms. Relational alignment allows all three ISO stakeholders to understand each other's domains, their processes and makes them comfortable to interact with each other. CVA capability generates conversations, increases collaboration and helps achieve common goals and decisions and enables the teams to work in a non-linear manner, as they understand each other's work processes. So as the developers work on specific

solution components of the system, the business side can be defining requirements for other components. Together these three dimensions keep the client and vendor working on the "same path" over the long term.

ISO Success Factors
Current IS research suggests that Information systems outsourcing success is an inter-firm outcome that is jointly driven by both client and vendor measures (Gottschalk and Solli-Saether, 2005). For this study, the list of eleven critical success factors for IT outsourcing includes: definition and needs management, resource exploitation across the alliance, cost reduction, relationship exploitation, vendor behavior control and stakeholder management (Gottschalk and Solli-Saether, 2005). These success factors represent a balanced set that does not simply focus on vendor side cost reductions and resource exploitation, but also includes client side factors like stakeholder management and needs definition. This collection of success factors form an inter-firm measure that highlights the importance of the impact of client vendor alignment on ISO outcomes

4. RESEARCH MODEL and HYPOTHESES

Outsourcing governance involves many operational and strategic decisions such as the definition and prioritization of IS projects, the funding and allocation of resources and measuring the value of such projects. Governance attempts to counteract the uncertainties posed by the increasingly complex and interconnected hosted technical environment. Since it is difficult to specify complete service level agreements (SLA) inside contracts, strict contractual governance or "mechanistic" governance is limited to outsourced systems that are "commodities" and are well understood and bounded in terms of their extensiveness and completeness and every detail and scenario and outcome is pre-specified in the contract (Goo, et.al.,2009). Under relational governance, the client and vendor can rely more on their ongoing relationship and mutual trust for deciding about emerging situations and managing the outsourcing arrangement, rather than following a contract very closely. Figure 1 shows the research model and hypotheses.

Building Client-Vendor Alignment Capability
Outsourcing governance typically falls into two categories – contractual and relational governance (Goo, et.al., 2009; Srivastava & Teo, 2012). Most outsourced work is fully or partially governed by contractual governance using a

---

formal contract between the client and vendor. Such client–vendor contracts describe the expected outcomes and behaviors of the work and can be tracked and measured per the vendor's performance.
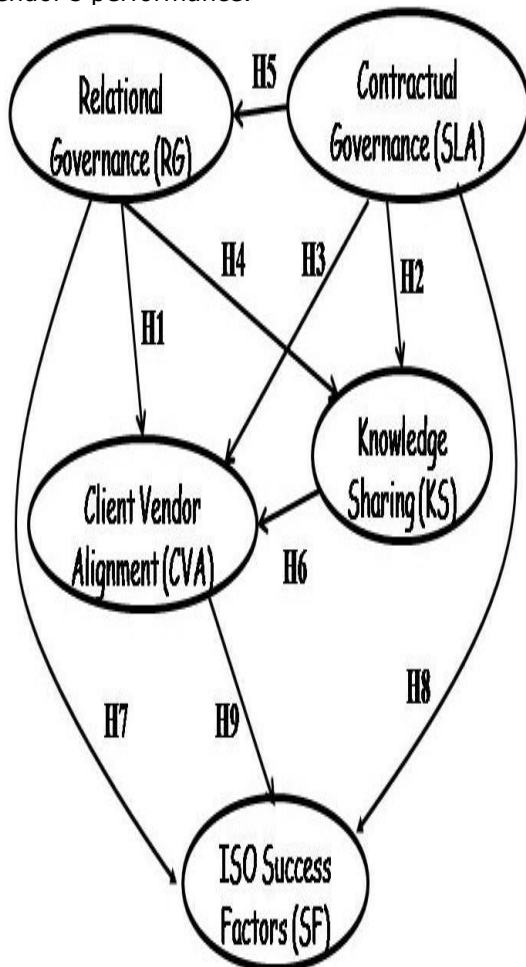

Figure 1: Research Model

Contractual governance and relational governance mechanisms allow the client and vendor to develop a common vision and establish a working structure. Trust enables the workers to work more cooperatively, limiting the power and positional rivalries. A stronger common identity fosters common goals among the workers and common norms enable members to transcend the diversities that are inherent in a multi-cultural organization and make communications smoother. These facets of relational governance can play a large part in the effectiveness and success of the outsourced processes, how much synergy is achieved between client and vendor personnel and the extent of tacit knowledge sharing (Inkpen and Tsang, 2005). By specifying relational governance elements – (1) staff feel safe to

explore and share new ideas without fear of failure, leading to better process execution (structural alignment), and (2) shared business vision is developed between client and vendor staff that establishes better strategic alignment.

Therefore, we have:

H1: Greater the Relational Governance Elements higher the level of Client-Vendor Alignment Capability.

H7: Greater the Relational Governance Elements higher the level of Strategic Outsourcing Success Factors.

An outsourcing contract provides a well defined framework in which client and vendor can understand each other's rights, duties, and responsibilities in the outsourcing arrangement (Goo, et.al., 2009). The contract also specifies policies and strategies underlying the arrangement. The contract enables firms to establish working relationships (relational alignment) and exchange knowledge about work processes (structural alignment) and share their long term vision (strategic alignment). Consequently we posit:

H3: Greater the Contractual Governance Elements higher the level of Client-Vendor Alignment Capability.

H8: Greater the Contractual Governance Elements higher thelevel of IS Outsourcing Success Factors.

Governance Elements support Knowledge Sharing
Outsourcing governance elements facilitate more cooperative, long-term exchange relationships between the client and vendor (Poopo and Zenger, 2002). Contractual governance elements document mutually agreed upon policies and procedures for dealing with dynamic situations during the outsourcing and lays the framework for knowledge exchanges (Goo, 2009). Likewise, relational elements of governance such as social capital and norms of relationships help close knowledge gaps in offshore ISO and serve as a lubricant for workers to get support and advice well beyond the organizational hierarchy or contracts, to enable them to share knowledge and get things done more effectively (Ghosh and Scott, 2009). Therefore, contractual governance and relational governance are needed for successful knowledge sharing (Palvia 2010).

H2: Greater the Contractual Governance Elements higher thelevel of Client-Vendor knowledge sharing.

H4: Greater the Relational Governance Elements higher the level of Client-Vendor knowledge sharing.

H5: Greater the Contractual Governance Elements higher the level of Relational Governance.

Knowledge Sharing builds Alignment Capability
The process of managing strategic ISO is often a "learning experience" in which the client may have to adapt and adjust the linkages that tightly couple the offshored activities with their internal skills and processes (Larsen, et.al., 2012). The client and vendor build inter firm organizational capabilities and structures by exchanging knowledge which enables the client to effectively exploit the vendor's resources and quickly address the uncertainties that are likely to be faced during the outsourcing period (Plugge, et.al., 2013). The knowledge sharing among client and vendor helps build and sustain the alignment capability by addressing emergent issues (Grant, 2003). We posit:

H6: Greater the Client Vendor Knowledge Sharing higher the level of Client Vendor alignment capability.

Alignment Capability supports Strategic ISO Success Factors
Both client and vendor develop and use internal resources to respond to the demands of the ISO and shifts in the business environment. Dynamic capabilities such as client-vendor alignment are particularly important to adapt to changing environments and achieve success over the long term in strategic ISO (Lee and Kim, 1999). The vendor needs to continuously make important decisions in order to improve its operational performance while supporting its clients' strategic goals with a long-term orientation. Developing and managing interfirm capabilities jointly with the vendor have been found to be keys to achieving greater outsourcing success for the client (Weigelt, 2013). When alignment capability is strong, the client provides the vendor with a unifying vision to enable the client to lead in their business and marketplaces and support the client's strategy across all business segments and stakeholder groups (Palvia, et.al., 2010). Therefore we have:

H9: Greater the Client-Vendor Alignment Capability higher the level of Outsourcing Success Factors.

5. DATA COLLECTION & ANALYSIS

A questionnaire (Table 1 of the Appendix) with multiple items (5 point Likert scale) for each construct was developed and pilot tested. After the pilot survey determined that the items meet content validity, the final survey was conducted as a convenience sample of key business and IT personnel from the client and vendor side of four strategic outsourcing deals. The four client companies are based in North America in the oil and gas exploration and energy production industry. The vendor is based out of India.

Characteristics of Client Organizations
The four client companies chosen for this study are medium sized energy exploration companies located in North America (USA and Canada). Some of the characteristics and details of these four firms (identified as A, B, C and D) are in Table 2 of the appendix. Each of these companies identified multiple information systems projects that were farmed out to a large ISO vendor based in India.

The North American energy exploration industry has recently experienced turbulent times with industry consolidation, labor shortage, government regulations, and economic conditions creating major fluctuations in commodity prices and reduction in consumer energy demand. Such environmental uncertainties are causing each of the firms to invest in new strategic information systems to improve various aspects of their business such as managing drilling sites and optimizing product extraction and distribution, improving capital equipment utilization and safety and training of their human capital. The size and public availability of geological data has allowed the vendor (India based) to build systems that can help these firms achieve operational efficiency. To achieve market focus and responsiveness, the firms needed to restructure their functional orientation around processes through organizational reengineering, updated infrastructure and technology use. However, the highly rigid and inbred organizational culture, strategy and relatively longstanding IS practices of the client had to be aligned with the vendor's system capabilities and implementation processes to achieve transformational results. A mix of contractual and relational governance elements were put in place to build client-vendor alignment capability and manage the outsourcing projects for the long term. Building client vendor

knowledge sharing and alignment capability was seen to be important for the long term success of the outsourcing deals.

Data Analysis

There were 107 completed surveys from the 200 surveys distributed for a response rate of 53%. Of these 107, 33 participants were from the four client firms and 74 from the vendor firm.　The demographics (Table 3 in Appendix) of the survey participants show an average of 5.49 years of experience on their current job, with 8.84 years of total professional experience and an average of 4.45 years of post high school education.　69 of the 107 were males and 38 were female.　The job titles of the survey participants included: business management, IT management, systems analysts, systems development and IT support. The breakdown of which client-vendor projects the participants identified with, are also provided in Table 3.

Smart PLS was used to test the measurement model for construct validity and reliability.　The results of the measurement model validity tests are listed in Table 4 (Appendix).　In order for the measurement model to be valid, the composite reliability of the reflective constructs are above 0.60 and the square root of the AVE measure of the construct is greater than the construct's correlation with other constructs.　Both these rules are satisfied for the three reflective constructs – Client-Vendor Alignment (CVA), Knowledge Sharing (KS) and Relational Governance (RG).　The two formative constructs – Success Factors (SF) and Contractual Governance (SLA) indicate adequate construct validity to continue with the Smart PLS analysis of the structural model and test hypotheses. The results of the hypotheses testing is shown in Table 5 (Appendix).

The hypothesis testing results indicate that all proposed hypotheses were supported by the survey data.　Contractual governance elements and relational governance elements both impact knowledge sharing (H2 and H4), the development of the client vendor alignment capability (H3 and H1) and support the ISO success factors (H7 and H8).　The provisions of the SLA used in contractual governance drive the level of relational governance (H5), as the frequencies and types of communication and cooperation between the client and vendor are stated in the SLA's (Service level agreements).　The level of knowledge sharing between the client and vendor staff drives the level of client vendor alignment capability (H6).　Finally, the level of the client vendor alignment capability impacts the ISO success factors (H9).

6. CONCLUSIONS

The results indicate that client vendor alignment capability is an important inter-firm capability between the client and vendor and allows the organizations to prioritize strategic project decisions and then operationalize those decisions using intersecting work structures and business processes of both the client and vendor firms.　As opposed to conventional short term outsourcing of localized IS projects, strategic ISO refers to the long term sourcing of Information systems that impact several enterprise business processes. The scope and requirements of such strategic projects have the potential to change over the course of the outsourcing deal.　For these types of long term and large projects, it is very difficult to hash out contractual terms in great detail and define work items exhaustively at the beginning. It is more promising over the long run in these outsourcing arrangements to establish a joint client-vendor project framework to support long term sourcing processes that will play a role in the definition and delivery of the Information System.　Such a framework involves building joint client vendor capabilities, which tackles strategic roles and responsibilities in the relationship as well as operational roles and responsibilities. The contribution of this research paper is the definition of such a capability, referred to as client vendor alignment capability. The paper defines and validates a measure of the client vendor alignment (CVA) capability and finds theoretical support for the need to establish client vendor alignment capability for successful strategic outsourcing engagements.　The study finds that establishing CVA in an outsourcing engagement requires the adoption of sophisticated vendor management activities that rely on elements of both contractual and relational governance (Willcocks, et.al.,1999). The paper also finds the importance of contractual governance mechanisms that help build the inter-firm relationship and in setting up inter-firm communications and knowledge sharing (Goo and Huang, 2008).　The study finds that knowledge sharing plays an important role in creating a better understanding of each firm's plans, objectives, resources and processes and building consensus on the opportunities and challenges facing the deal (Klein and Rai, 2009).

It is critical for the success of an offshoring strategy to bring consensus in all levels of the organization. As information systems outsourcing (ISO) engagements become bigger value and

span longer terms, clients also seek greater value and diverse objectives (Mukherjee, et.al., 2013).In this scenario, client vendor alignment becomes an important capability for the long run.

Future Research

The results of this research present a strong case to conduct a larger multi-industry study of strategic outsourcing practices. There is a need to understand the content of strategic outsourcing deals such as infrastructure, internal applications or customer facing applications. Additional demographic questions can be added to collect information about the contents of outsourcing deals and if certain dimensions of alignment play a more important role for different deals. A mixed approach can also be adopted in the future as the research of strategic outsourcing and client vendor alignment is relatively new and less established in the IS literature. A qualitative study using interviews and/or surveys with open ended questions can also be used to collect case data for definition of the constructs and discovery of the construct relationships and research model using grounded theory. A follow-on survey can then be used to collect quantitative data to measure the constructs and test the relationships induced in the research model.

The profile of the oil and gas industry is unique, as it faces a diminishing labor pool, volatility in raw materials input prices and output retail energy prices along with the strict government regulations that place a significant compliance burden on the industry. Such an environment forces tightening of business margins and forces decision making under greater stress. The oil and gas industry also has cutthroat competition and frequent mergers and acquisitions that can create difficulties in information systems projects. Energy exploration is a capital intensive business as large amounts of money need to be invested to locate and develop energy resources. Because the oil and gas industry has been slower to change their business practices it is likely that the client firms are more likely to adopt the newer vendor suggested systems and development practices thus achieving alignment (CVA) more easily.

7. REFERENCES

Alavi, M. and Leidner, D. (2001). Review: Knowledge Management and Knowledge Management Systems : Conceptual Foundations and Research Issues, *MIS Quarterly* 25(1), 107-136..

Aubert, B., Patry, M. and Rivard, S. (1999). Managing the Risk of IT Outsourcing. *Proceedings of 32nd HICSS*.

Barney, J. B., Wright, M. and Ketchen, D.J. (2001). The resource based view of the firm: Ten years after 1991. *Journal of Management*, 27(3), 625-641.

Bharadwaj, A. (2001). A Resource-based Perspective on Information Technology Capability and Firm Performance: An Empirical Investigation. *MIS Quarterly*, 24(1), 169-196.

Carlile, P.R. (2004), Transferring, Translating and Transforming: An Integrative Framework for Managing Knowledge Across Boundaries, *Organization Science* 15(5), 555-568.

Chan, Y. (2002). Why Haven't We Mastered Alignment? The Importance of the Informal Organizational Structure.*MIS Quarterly Executive* 1(2), 97-112.

Chang, Y.B. and Gurbaxani, V. (2012). The Impact of IT-Related Spillovers on Long Run Productivity: An Empirical Analysis. *Information Systems Research* 23(3), 868-886.

Davenport, T.H., De Long, D.W. and Beers, M.C (1998). Successful Knowledge Management Projects, *Sloan Management Review* 40(1), 43-57.

Dibbern, J., Goles, T., Hirschheim, R. and Jayatilaka, B. (2004). Information Systems Outsourcing: A Survey and Analysis of the Literature. *The DATA BASE for Advances in Information Systems*, 34(4), 6-102.

Earl, M.J. (1996). The Risks of Outsourcing IT.Sloan *Management Review*, 37(3). 26-32.

Feeny, D.F. and Willcocks, L.P. (1998). Core IS Capabilities for Exploiting Information Technology. *Sloan Management Review, 40(1),* 9-21.

Flinders, K. (2014). Lufthansa signs long-term outsourcing deal with IBM, ComputerWeekly.com, accessed at: http://www.computerweekly.com/news/2240233142/Lufthansa-signs-long-term-outsourcing-deal-with-IBM.

Foogooa, R. (2008). IS Outsourcing – a strategic perspective.Business Process *Management Journal*, 14(6), 858-864.

Ghosh, B. and Scott, J.E. (2009). Relational Alignment in Offshore IS Outsourcing. *MIS Quarterly Executive*, 8(1), 19-29.

Goo, J. (2010). Structure of service level agreements (SLA) in IT outsourcing: The construct and its Measurement, *Information Systems Frontiers* (12), 185-205.

Goo, J. and Huang, C.D. (2008). Facilitating relational governance through service level agreements in IT outsourcing: An application of the commitment-trust theory, *Decision Support Systems*, (46), 216-232.

Goo, J., Kishore, R., Rao, H.R. and Nam, K. (2009). The Role of Service Level Agreements in Relational Management of Information Technology Outsourcing: An Empirical Study. *MIS Quarterly*, 33(1), 119-145.

Gonzalez, R., Gasco, J. &Llopis, J. (2010), Information Systems Offshore Outsourcing: An Exploratory Study of Motivations and Risks in Large Spanish Firm. *Information Systems Management*, 27, 340-355.

Gottschalk, P. and Solli-Saether, H. (2005). Critical Success Factors from IT Outsourcing Theories: an Empirical study. *Industrial Management and Data Systems*, 105(6). 685-702.

Grant, G.G. (2003). Strategic Alignment and Enterprise Systems implementation: the case of Metalco. *Journal of Information Technology,* 18, 159-175.

Greaver, M.F. (1999). Strategic Outsourcing: A Structured Approach to Outsourcing Decisions and Initiatives. New York, AMACOM.

Inkpen, A.C. and Tsang, E.W.K. (2005). Social Capital, Networks and Knowledge Transfer. *Academy of Management Review,* 30(1), 146-165.

Kern, T. and Willcocks, L.P. (2000). Exploring Information Technology Outsourcing relationships: Theory and practice. Journal of Strategic Information Systems, 9(4), 321-350.

Kishore, R, Rao, H.R., Nam, K., Rajagopalan, S. &Chaudhury, A. (2003). A Relationship Perspective on IT Outsourcing.*Communications of the ACM*, 46(12), 87-92.

Klein, R. and Rai, A. (2009). Interfirm Strategic Information Flows in Logistics Supply Chain Relationships, *MIS Quarterly*, 33(4), 735-762.

Ko, D. Kirsch, L. and King, W.R. (2005). Antecendents of knowledge transfer from consultants to clients in enterprise system implementations, *MIS Quarterly*, 29(1), 59-85.

Lacity, M.C. and Hirschheim, R. (1993), Information systems outsourcing bandwagon.*Sloan Management Review*, 35(1), 73-86.

Lacity, M.C. and Willcocks, L.P. (1998), An Empirical Investigation of Information Technology Sourcing Pracxtices: lessons from Experience, *MIS Quarterly*, 22(2), 363-408.

Larsen,M.M., Manning, S., Pedersen, T. (2012). Uncovering the hidden costs of offshoring: the interplay of complexity, organizational design and experience. *Strategic Management Journal*

Lavie, D. (2006). The competitive Advantage of interconnected firms: An extension of the Resource Based View. *Academy of Management Review*, 31(3), 638-658.

Lee, J. N and Kim, Y. G. (1999). Effect of Partnership Quality on IS Outsourcing Success. *Journal of Management Information Systems*, 15(4), 29-61.

Luftman, J. and Brier, T. (1999). Achieving and Sustaining Business-IT Alignment.*California Management Review*, 42(1). 109-122.

Mukherjee, D., Gaur, A.S. and Dutta, A. (2013). Creating value through offshore outsourcing: An integrative framework. *Journal of International Management* (19), 377-389.

Palvia, P.C., King, R.C., Xin, W. and Palvia, S.C.J. (2010). Capability, Quality and Performance of Offshore IS Vendors: A Theoretical Framework and Empirical Investigation. *Decision Sciences*, 41(2), 231-270.

Phillips, P. (2014). Long-term outsourcing deals defy market trend, ComputerWeekly.com, Accessed at http://www.computerweekly.com/feature/Long-term-outsourcing-deals-defy-market-trend.

Plugge, A., Bouwman, H. and Molina-Castillo, F.J. (2013). Outsourcing capabilities, organizational structure and performance quality monitoring: Toward a fit model. *Information and Management,*(50) 275-284.

Poppo, L. and Zenger, T. (2002). Do Formal Contracts and Relational Governance Function as Substitutes or Complements? *Strategic Management Journal,* 23(8). 707-725.

Ross, J. W.; Beath, C. M.; and Goodhue, D. L. (1996). Developing Long-term Competitiveness through IT Resources. *Sloan Management Review*, 38(1), 31-45.

Rai, A., Pavlou, P.A., Im, G. and Du, S. (2012), Interfirm IT Capability Profiles and Communicaions for Cocreating Relational Value: Evidence from the Logistics Industry, *MIS Quarterly.*36(1), 233-262.

Reich, B. and Benbasat, I. (2000), factors that Influence the Social Dimension of Alignment between Business and Information Technology, *MIS Quarterly* 24(1), 331-334.

Rottman, J.W. and Lacity, M.C. (2004). Twenty Practices for Offshore Sourcing.*MIS Quarterly Executive* 3(3), 117-130.

Sabherwal, R. (1999). The Role of Trust in Outsourced IS Development Project. *Communications of the ACM*, 42(2), 80-86.

Srivastava, S.C. and Teo, T.S.H. (2012). Contract Performance in Offshore Systems Development: Role of Control Mechanisms. *Journal of Management Information Systems*, 29(1). 115-158.

Tanriverdi, H. (2005). Information Technology Relatedness, Knowledge Management Capability, and Performance of Multibusiness Firms. *MIS Quarterly*, 29(2), 331-334.

Willcocks, L.P., Lacity, M.C. and Kern, T. (1999). Risk mitigation in IT outsourcing strategy revisited: longitudinal case research at LISA. *Journal of Strategic Information Systems*, 8(3), 285-314.

Willcocks, L.P. and Kern, T. (1998). IT Outsourcing as Strategic Partnering: The Case of the UK Inland Revenue. *European Journal of Information Systems*, 5(1), 29-45.

Willcocks, L.P., Hindle, J., Feeny, D. and Lacity, M.C. (2004). IT and Business Process Outsourcing: The Knowledge Potential. *Information Systems Management*, 21(3), 7-15

Wade, M. and Hulland, J. (2004). Review: The Resource Based View and Information Systems Research: review, Extension and Suggestions for Future Research. *MIS Quarterly*, 28(1),107-142.

Weigelt, C. (2013). Leveraging Supplier Capabilities: The role of locus of capability deployment. *Strategic Management Journal,* 34, 1-21.

Appendices and Annexures

Table 1: Survey Items

Years of Job Expr. _____Years of Prof Expr _____ Job Title: _____ Years of College/University (post K-12) _____Gender : Male _____ Female _____	
Relational Governance (RG)	We have good teamwork among staff in the ISO relationship. We express diverse/conflicting views among staff in the ISO relationship. Client and vendor staff share common goals and mutual understanding.
Contractual Governance (SLA)	Service level agreements (SLA) clearly define scope and objectives of the ISO. SLA defines the ownership of processes and the measurement of their outcome. SLA have provisions of Communication and norms of behavior among staff. SLA has enforcement for the development and deployment of IS applications.
Knowledge Sharing (KS)	Business knowledge is freely exchanged between client and vendor. System and technical knowledge is freely exchanged between client and vendor. We have multiple channels of knowledge sharing - synchronous and asynchronous.
Client Vendor Alignment (CVA)	We jointly make IT Needs decisions and application prioritizations. There is fit between the priorities and activities of client and vendor. Our operational processes support joint work on projects.
ISO Success Factors (SF)	We can successfully Define and manage IT needs. We exploit a mix of resources from client and vendor with division of labor. We can successfully reduce complexity and uncertainty in IT tasks. We avoid opportunistic behavior from either client or vendor. We manage costs efficiently and successfully support all stakeholders.

Table 2: Characteristics of Four Client Firms participating in this Study

ID	Primary Business Activities	Strategic Business Goals		Outsourced Strategic Information Systems Projects	Major Information Systems Project Challenges
A	Exploration, Extraction and Distribution Logistics	1.	Optimize equipment uptime and utilization	Big data Integration for geological exploration, Business Analytics for process management and equipment maintenance, End to end tracking of extracted materials	Systems Integration; Resolve conflicts between internal and external systems; Business Process reengineering
		2.	Reuse of external geological data		
B	Exploration, Extraction and Distribution Logistics	1.	Real time visibility/managementofField Operations	Design, deploy and use of sensors and unstructured data in operational dashboards; Collaboration and knowledge management systems	Lack of collaboration between business leaders and IT leadership; Fragmented processes; Poor change management of systems deployment
		2.	Quick Evaluation of Extracted Assets		
		3.	Increase Collaboration		
C	Exploration	1.	Optimize allocation of Investment capital to most productive sites	Design, deploy and use of sensors and unstructured data in operational dashboards to monitor exploration sites	Requirements collection and translating them into real project phases; Lack of IT project management, working protocols and procedures
		2.	Optimize the exploration efficiency from these sites		
D	Production and Distribution	1.	Integration of supply chain from production sites to distribution sites	New production management system, Integration with Logistics/warehousing systems of distributors; Collection and aggregation of big data for market forecasts	Lack of Project Cost benefit analysis capabilities; Week Project Leadership and lack of project sponsors
		2.	Improve product distribution channels		

Table 3: Demographics of Survey Participants

	(Total: 107 Respondents)	
	Mean	**StdDev**
Years of Schooling (include 13 years for K-12)	17.45	2.46
Years on Current Job	5.49	2.35
Years in Profession	8.84	4.48
Gender	Males: 69 Females:38	
Outsourcing Arrangement Identified	A : 23, B : 39, C :17, D : 28	

Table 4: Measurement Model Construct and Validity Measures & Correlations

Construct	Sq. RootAVE	Composite Reliability	R-square OR Communality(*)	Cronbach Alpha	Construct Correlations				
					CVA	KS	SF	RG	SLA
CVA	0.753	0.793	0.653	0.626	1.000	0.000			
KS	0.891	0.919	0.465	0.868	0.667	1.000			
SF			0.421*		0.637	0.458	1.000		
RG	0.901	0.928	0.442	0.883	0.782	0.662	0.524	1.000	
SLA			0.3503*		0.580	0.562	0.299	0.665	1.000

Table 5: Results of Hypotheses Testing

Hypothesis	Path Coeff.	Std. Dev.	T-Value	Hypothesis Supported
H1: Relational Governance (RG) -> Client Vendor Alignment (CVA)	0.6945	0.080	8.909	YES
H2: Contractual Governance (SLA) -> Knowledge Sharing (KS)	0.5701	0.053	10.693	YES
H3: Contractual Governance (SLA) -> Client Vendor Alignment (CVA)	0.5986	0.057	10.136	YES
H4: Relational Governance (RG) -> Knowledge Sharing (KS)	0.5204	0.093	5.5787	YES
H5: Contractual Governance (SLA) -> Relational Governance (RG)	0.6739	0.049	13.413	YES
H6: Knowledge Sharing (KS) -> Client Vendor Alignment (CVA)	0.2472	0.071	3.581	YES
H7: Relational Governance (RG) -> ISO Success Factors (SF)	0.5849	0.131	4.450	YES
H8: Contractual Governance (SLA) -> ISO Success Factors (SF)	0.3240	0.121	2.468	YES
H9: Client Vendor Alignment (CVA) -> ISO Success Factors (SF)	0.6228	0.173	3.531	YES

Malvertising - A Rising Threat
To The Online Ecosystem

Catherine Dwyer
cdwyer@pace.edu

Ameet Kanguri
ak23433n@pace.edu

Seidenberg School of Computer Science & Information Systems
Pace University
New York, New York, USA

Abstract

Online advertising is a multi-billion dollar industry that supports web content providers around the globe. A sophisticated technology known as real time bidding (RTB) dominates the advertising landscape, connecting advertisers with specific online customers of interest. With RTB, when web visitors connect to a site, advertising networks are notified of space available on that site along with what can be gleaned about the visitor. These combinations of space and visitor are auctioned, and the winning bid's ad content is served to the web visitor. The entire process, from a visitor landing on a publisher's page to ads being auctioned, selected and served, takes 200 milliseconds, the time needed to snap your fingers. This tightly choreographed interaction is a technical marvel, but one with built in risks. The just-in-time collaboration between ever changing technology providers gives an opening to malicious actors, who through devious means, use ad networks to deliver malware rather than ads. Delivering malware as an ad is called malvertising, and its presence on otherwise credible sites is dangerous, undermining the business models of trustworthy publishers and legitimate online advertisers. The purpose of this paper is to introduce malvertising, describe its relationship with online advertising, and identify the risks RTB and malvertising bring to the online ecosystem.

Keywords: malware detection, malvertising, online advertising, ad blockers, real time bidding (RTB).

1. INTRODUCTION

The term malvertising is constructed by combining "malware" and "advertising". According to the SANS institute, malvertising is "the installation of unwanted or outright malicious software through the use of internet advertising media networks, exchanges and other user supplied content publishing services common to the Social Networking space," (Salusky, 2007).

What makes malvertising a special threat to the Internet? Malvertising cleverly uses the power of targeted advertising to specifically deliver malware to victims who visit trusted sites such as forbes.com (Patrizio, 2016), Spotify (Hern, 2016) the BBC or The New York Times (Mihalcik, 2016). By using online advertising tools to target victims of interest via algorithm, for example employees in the defense industry (Invincea, 2015a), and unsuspecting visitors to trusted sites (Ducklin, 2016), malvertising can upend the most important economic driver for the Internet -- advertising revenue -- and damage the reputation of well know sites.

The web as we know is it funded in large part by advertising revenue. Most online content providers, with very few exceptions, earn the bulk of their revenue from digital advertising, with

little or no revenue from subscriptions or fees for site access (Deloitte, 2016). The world's largest online companies, Google and Facebook, derive most of their revenue from online advertising in one form or another (Gjorgievska, 2016).

2. THE NATURE OF ONLINE ADVERTISING

The online advertising ecosystem is an multifaceted technical network matching buyers and sellers of ad space on pages currently under view by web visitors who match specific profiles of interests. Given this happens on millions of web pages seen by millions of web visitors, all within a window of 200 milliseconds (Lederer, 2014), online advertising can be considered one of the most technologically advanced information systems ever developed.

Several recent technology drivers intersect in the delivery of online ads. One is the data collection and profiling of consumers based on their social, mobile and online activities. Based on the collection of terra-bytes of data, companies such as Experian have developed specific customer profiles. Experian has identified 19 categories and 71 sub-categories of consumer profiles. These include the category called "Singles and Starters," and its six sub-categories, including "Digital Dependents," "Colleges and Cafes," and "Striving Single Scene," (Experian, 2014).

The second technology driver is the development of real time bidding (RTB) systems. Advertisers have always been interested in finding the right audiences for their products. RTB automates this process. "RTB helps media buyers find audiences at scale," according to a Google white paper on RTB (Google, 2011).

Google introduced bidding for ads associated with specific search terms with AdWords in 2000 (Mehta, Saberi, Vazirani, & Vazirani, 2007). With AdWords, advertisers could compete with each other to serve ads to users based on search terms and cookie data.

The next stage in the development of RTB was to expand this bidding and audience targeting system to other domains, such as display and banner advertising. Companies such as RightMedia and DoubleClick expanded the bidding process beyond search advertising. By 2011 RTB had become the dominant mechanism for online advertising (Chen, Berkhin, Anderson, & Devanur, 2011).

As of 2016, 23 different sub-categories of companies have been identified that participate in the market for online display advertising (Kawaja,

2016). For the purposes of this paper, we will focus on these players:

1) Publisher - Companies or individuals that generate content for consumption by consumers. Publishers monetize their content by putting up ads besides their content. Examples of publishers include NYTimes.com and Forbes.com.

2) Supply Side Platform (SSP) - A supply-side platform or sell-side platform (SSP) is a technology platform hired by publishers to manage their online advertising space inventory, fill it with ads, and receive revenue. Examples of SSPs include Rubicon and Pubmatic.

3) Demand Side Platform/Ad network (DSP) - A demand side platform is hired by advertisers to manage its bids for online ad space. Examples of DSPs include MediaMath and InviteMedia.

4) Ad exchange – Like a stock exchange, it brings together buyers and sellers of online ad space. Examples of ad exchanges include DoubleClick (owned by Google) and OpenEx.

5) Digital marketer - Advertising agencies representing large companies wanting to post advertisements online. Examples include OmniconGroup and WPP (Ju, 2013).

The interaction that takes place in online advertising is diagrammed in Figure 1 (Kneen, 2015). When a web visitor lands on a web page (labeled as step 1), the page is loaded along with an ad tag embedded acting as a placeholder (step 2). This tag triggers a further call to an SSP, passing along the ad dimensions and the identity of the publisher (step 3 and 4). From there the SSP reads the SSP cookie (step 5) from the user's machine (most users already have a SSP cookie which is created while visiting an earlier site). Major SSPs claim to have cookie coverage of 80% across US users (Ad Ops Insider, 2010).

The SSP then requests bids through the ad exchange from a host of DSPs (step 6 and 7). The SSP cookie is passed on to each DSP and this helps the DSPs value the impression. The DSP matches the cookie data to their own cookie data (step 8, 9 and 10), which in-turn is tied to a huge cache of marketer data and third party data. In a nutshell this data is a detailed browsing history of the user that marketers and data brokers have collected. The richer the data available about the user, the higher the bids from DSPs (Ad Ops Insider, 2010).

Using this information the DSPs place bids and send an ad redirect link to the SSP in case it wins the bid. The SSP selects the winning bid, and sends the DSP link to the user, whose browser

then calls the marketer's server to display the ad (steps 11 and 12). The RTB ad serving process is complete. The entire process takes about 200 milliseconds (Kneen, 2015).

3. MALVERTISING AND RTB

Malvertising is the seeding of malicious code in online advertisements and delivering these to unsuspecting users visiting common and trusted websites, such as huffingtonpost.com, twitter.com, and cnn.com (Mimoso, 2015).

The nature of the online advertising ecosystem and the rapidly changing collection of companies participating in online advertising has created an opportunity for malicious actors to masquerade as advertisers (Zarras et al., 2014), who can use the RTB advertising ecosystem to quite effectively deliver malware (Segura, 2015), and even specifically target individuals of interest, such as those that work in defense industries (Invincea, 2015a).

An example of targeted exploits delivered via an advertising network is the 2014 hack of the Reuters site, specifically news articles about Syria. If a news article about Syria was selected, the visitor was then redirected to a web page on the hacker's website (see Figure 2). The attack was fine-tuned to ignore most of the website and only act with specific content. This was achieved via an advertising network named Taboola that managed display ads on the website. Through targeting, the attackers could determine information on who was reading a specific article on the website and target only those users (Jacobs, 2014).

Cyphort Labs, a provider of anti-malware services, issued a report that noted an increase in documented malvertising campaigns of 325% (2015). For example, MalwareBytes has documented the presence of malvertising on msn.com (Segura, 2016).

Online malware is a serious problem, one that affects individuals and organizations. An important element of safe internet use is avoiding suspicious, criminal, or inappropriate websites ("Safe Internet Use," 2016). Another important practice is vigilance with email, and staying away from links that seem suspicious in any way ("Spam & Phishing," 2016).

It certainly is a safer practice to only visit legitimate sites, those whose authenticity can be independently verified. While this is excellent advice, the use of online advertising networks by malicious actors to distribute malware on legitimate sites means that more rigorous methods must be developed to control the distribution of malware on the Internet.

Most sites and publishers rely heavily on online advertisements to monetize visits to their sites. According to the Interactive Advertising Bureau (IAB), online advertising in the USA reached $27.5 billion in the first half of 2015, a 19% rise over first half of 2014 (IAB, 2015). It is expected to continue to grow at a similar pace over the next few years.

RTB is a sophisticated technological interchange that has created a marketplace where many technology companies exchange bids and serve ads. The multi-party nature of this highly automated bidding exchange has introduced a risk in the form of malvertising.

Publishers are connected with advertisers by a network of companies, and the entire process is opaque to the end user. Ads are sold via a bidding process, and apart from the type of ad displayed, the publisher does not control which advertiser wins the bid and post ads. This allows not just legitimate parties but also miscreants to bid for ads (Invincea, 2015a).

Attack methods delivered through malvertising include deceptive downloads, link hijacking, and drive by downloads. Deceptive downloads lure their victims to download malicious software components disguised as browser plugins and other software add-ons. This happens by having the user believe that to access some desirable content they need to install a particular software component.

In link hijacking the user is surreptitiously redirected away from safe websites to sites with exploits. This is done by inserting malicious code in the ads that causes the redirect.

The most dangerous method is called a "drive-by-downloads". The risk from drive by downloads is that the user may infect his or her computer by merely visiting the website, even without directly interacting with malicious part of the page. In this scenario the malicious exploit originates from the ad network server and probes for browser vulnerabilities. The most common targets among attackers are machines with outdated plugins for Java and Flash (Zarras et al., 2014).

Malvertising is the use of online advertising as a vector to deliver malware. It involves the injection of malicious or malware laden advertisements into legitimate, recognized web sites such as Yahoo.com (Grandoni, 2015), MSN.com (Segura, 2016), and dictionary.com (Invincea, 2015b). By injecting malware via advertising into high profile web sites, users not

typically vulnerable to malware can be targeted. This infection can take place "silently," through techniques such as drive by downloads that do not require any action by the web site visitor other than opening the page in a browser.

A report by the IAB and Ernst and Young included this sobering comment about malvertising: "the need to click on the malware to be infected is a common misconception of the public," ("What Is An Untrustworthy Supply Chain Costing The U.S. Digital Advertising Industry?," 2015). Through malvertising, the profiling capabilities of online advertising can be re-purposed to target individuals and organizations of interest, for the distribution of ransomware, and theft of intellectual property.

The security firm Invincea has documented dozens of these attacks taking place on sites such as cbssports.com, match.com, answers.com, and realtor.com (Invincea, 2015b).

4. MALVERTISING AND AD BLOCKERS

If malware can be delivered through advertising networks, then it has been suggested that using an ad blocker will also block malvertising. In 2015 Edward Snowden endorsed the use of ad blockers to protect against attacks through malvertising, saying "as long as service providers are serving ads with active content that require the use of Javascript to display, that have some kind of active content like Flash embedded in it, anything that can be a vector for attack in your web browser — you should be actively trying to block these," (Lee, 2015). While many claim that ad blockers can protect you, no empirical studies have been published to date that prove that ad blockers protect against malvertising.

Ad blockers have been at the center of a dispute between publishers and the developers of ad blocking software. The head of the IAB has criticized ad blockers, and the organization has begun a public campaign against them, arguing they "are stealing from publishers, subverting freedom of the press, operating a business model predicated on censorship of content and ultimately forcing consumers to pay more money for less—and less diverse—information." (Heine, 2016). Some publishers prevent web visitors using ad blockers from viewing content, including wired.com and forbes.com (Schneier, 2016).

The use of ad blockers by online users has been criticized by publishers. Ad blockers are found on 15% of all US internet browsers ("The 2015 Ad Blocking Report," 2015). Most ad blockers are installed as browser plugins, with the two most popular versions being AdBlock and AdBlock plus.

Irrespective of the ad blocker used, most ad blockers rely on a collaborative database called EasyList ("AD BLOCKERS a guidebook for publishers, advertisers and Internet users," 2014) . EasyList gathers a list of regular expressions that recognize an ad versus other content. These are sequences of code written to spot keywords or frameworks inside a webpage. Contributors submit any new sequences to the community who then reviews and approves it. Having more than 80,000 expressions it is largest reference database for all ad blockers.

Ad blockers do not differentiate between legitimate ads and malvertising, they block both. If the expression of code pattern is found on the web page the ad is blocked. This acts like a double-edged sword. While on one side with an updated database and a vibrant community adblockers block most malware, they also block legitimate ad content that is displayed on websites. But with ad blockers hurting earnings of publishers, a few of them have resorted to not displaying their content (or charging a fee) if they detect an ad blocker installed on the browser. Forbes (Patrizio, 2016) and Wired (Zorabedian, 2016) are more recent publishers who do not allow those using an ad blocker to view content for free on their site.

5. RISKS TO THE ONLINE ECOSYSTEM

The more automated online advertising is, the greater the efficiencies built into the system, the greater the opportunity for a malicious actor to exploit RTB.

There are challenges for publishers and online advertisers that make it more difficult to address the risks of malvertising and RTB. For one, publishers do not make as much money from online content as they made with print versions in the past and are vulnerable to any disruption in online revenue.

Secondly, online advertising depends on speed. One technique to disrupt malvertising is to place stricter controls over what files can be served as ads, however this can only slow the process down. The actual ad content does not come from either the publisher or the ad exchange, it comes from a separate technology company that optimizes its delivery. So there is a security supply chain problem in place. Checking the validity of ad content will only make the process less efficient and more time consuming.

The proliferation of malvertising on trusted sites has led businesses to turn to security solutions such Blue Coat that maintains a blacklist of known

malware sites, including a number of ad networks. This acts like a super ad blocker, blocking any ad delivery to a corporate environment (Mimoso, 2015).

For high income consumers visiting trusted sites like Forbes.com, they are attractive bait for exploits such as ransomware delivered through advertising. The success of these exploits are directly related to RTB, says Pat Belcher, director of malware analysis at the security company Invincea. "RTB has made it easier for malware authors to target individuals. Before RTB, you had to compromise the ad delivery network. Now, you not only win bids and place ads, you can use the same platform to pinpoint and target anyone you want" (Mimoso, 2015).

In some ways, this dilemma resembles the troubles advertisers and publishers have encountered with the collection of web browsing data. It is the use of these vast troves of data to serve carefully targeted ads that raises privacy concerns, and trying to make a perfect match instantly, millions of times a day, has created an opening for malvertising that could undermine the trust that is the foundation of ecommerce and the online market.

In addition to the risk of malvertising, because ad bids are higher if more can be discovered about the digital profile of a web visitor (Ad Ops Insider, 2010), there is a perverse incentive for publishers to collect and share as much information as possible with ad networks. And ad networks then collaborate through cookie sharing to precisely identify who is the online viewer, whether that person is at work, at home using a tablet, or on the go using their smart phone (Schiff, 2016).

6. CONCLUSIONS

Computer security best practices encourage end users to deploy strong passwords and avoid suspicious links. These however do not protect against drive-by downloads delivered by malvertising. If you do have a strong password and do avoid suspicious links, what else do you need to do to avoid malvertising? It is critically important to keep browsers and all plug-ins updated. It has also been suggested that ad-blockers can also protect the end user from infection by malware, since the online ad is the vector of delivery for the malware, since the ad-blocker blocks the ad, in theory it also blocks the malware.

Right now, the web depends on advertising for most of its financial support. However, that business model has opened the door to malware attacks using online ads as a vector. While publishers can say that the use of ad blockers does hurt their revenue, is also means publishers have an obligation to protect their site from malvertising. Given that RTB depends on a window of 200 milliseconds to deliver an ad (Lederer, 2014), there needs to be another control mechanism to ensure that bad actors cannot exploit this bidding process to serve malware.

Online advertising has grown into a multi-billion dollar industry by allowing advertisers to serve ads based on individual profiles, geolocation, client machine, and even a specific range of IP addresses. These precise targeting capabilities also make malvertising an attractive option for malicious actors. The customized delivery of ads also allows malvertising to hide from detection by employing stealthy targeting schemes that alternate the placement of benign advertising with the sporadic placement of malware (Cyphort, 2015).

Combatting malvertising will require an intricate multi-platform effort. It will require vigilance and adoption of best practices by multiple actors, including publishers/web hosting sites, ad networks, and web surfers. Publishers must require ad networks to develop an active prevention plan in place against malvertising. And ad networks will need to be more vigilant about the content of the ads they serve. As online ads take on more dynamic properties, including embedded scripts that customize the ad's content and appearance, then ad networks will need strict controls to ensure those scripts do not inject malware. Web surfers must protect themselves by keeping their browsers up to date, and where possible, disabling vulnerable plugins such as Java and Flash. So it is up to publishers, online advertisers, and the people who use those sites to work together to ensure the security of the web.

7. REFERENCES

The 2015 Ad Blocking Report. (2015). Retrieved from https://blog.pagefair.com/2015/ad-blocking-report/

AD BLOCKERS a guidebook for publishers, advertisers and Internet users. (2014). Retrieved from http://www.secretmedia.com/whitepaper/adblocker_whitepaper.php

Chen, Y., Berkhin, P., Anderson, B., & Devanur, N. R. (2011). *Real-time bidding algorithms for performance-based display ad allocation.*

Paper presented at the Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining.

Cyphort. (2015). The Rise of Malvertising. Retrieved from http://go.cyphort.com/Malvertising-Report-15-Page.html

Deloitte. (2016). The impact of web traffic on revenues of traditional newspaper publishers. Retrieved from https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/technology-media-telecommunications/deloitte-uk-impact-of-web-traffic-on-newspaper-revenues-2016.pdf

Ducklin, P. (2016). Malvertising – When trusted websites go rogue. Retrieved from https://nakedsecurity.sophos.com/2016/03/16/malvertising-when-trusted-websites-go-rogue-security-sos-week/

Experian. (2014). Mosaic USA consumer lifestyle segmentation. Retrieved from http://www.experian.com/marketing-services/consumer-segmentation.html

Google. (2011). The Arrival of Real Time Bidding. Retrieved from http://static.googleusercontent.com/media/www.google.fr/en/en/doubleclick/pdfs/Google-White-Paper-The-Arrival-of-Real-Time-Bidding-July-2011.pdf

Grandoni, D. (2015). Hackers Exploit 'Flash' Vulnerability in Yahoo Ads. Retrieved from http://bits.blogs.nytimes.com/2015/08/03/hackers-exploit-flash-vulnerability-in-yahoo-ads/?smprod=nytcore-iphone&smid=nytcore-iphone-share&_r=0

Heine, C. (2016). IAB Chief Blasts Adblock Plus as an 'Immoral, Mendacious Coven of Techie Wannabes'. *adweek*. Retrieved from http://www.adweek.com/news/technology/iab-chief-blasts-adblock-plus-immoral-mendacious-coven-techie-wannabes-169194

Hern, A. (2016). Spotify hit by 'malvertising' in app. Retrieved from https://www.theguardian.com/technology/2016/oct/06/spotify-hit-by-malvertising-in-app

IAB. (2015). Digital Ad Revenues Surge 19%, Climbing to $27.5 Billion in First Half Of 2015. Retrieved from http://www.iab.com/news/digital-ad-revenues-surge-19-climbing-to-27-5-billion-in-first-half-of-2015-according-to-iab-internet-advertising-revenue-report/

Invincea. (2015a). A case study in successfully defeating malvertising attacks. Retrieved from https://www.invincea.com/2015/09/white-paper-a-case-study-in-successfully-defeating-malvertising-attacks/

Invincea. (2015b). Fessleak: The Zero-Day Driven Advanced RansomWare Malvertising Campaign. Retrieved from https://www.invincea.com/2015/02/fessleak-the-zero-day-driven-advanced-ransomware-malvertising-campaign/

Jacobs, F. (2014). How Reuters got compromised by the Syrian Electronic Army. Retrieved from https://medium.com/@FredericJacobs/the-reuters-compromise-by-the-syrian-electronic-army-6bf570e1a85b -.d54x8pbr8

Ju, R. (2013). Online Advertising Explained: DMPs, SSPs, DSPs and RTB. Retrieved from http://www.kbridge.org/en/online-advertising-explained-dmps-ssps-dsps-and-rtb/

Kawaja, T. (2016). Display LUMAscape. Retrieved from http://www.lumapartners.com/lumascapes/display-ad-tech-lumascape/

Kneen, B. (2015). HOW REAL TIME BIDDING, DSPS, SSPS, AND AD EXCHANGES WORK. Retrieved from http://www.adopsinsider.com/ad-serving/how-dsps-ssps-and-ad-exchanges-work/

Lederer, B. (2014). 200 Milliseconds: Life of a Programmatic RTB Ad Impression. *Programmatic Insider*. Retrieved from http://www.mediapost.com/publications/article/225808/200-milliseconds-life-of-a-programmatic-rtb-ad-im.html

Lee, M. (2015). EDWARD SNOWDEN EXPLAINS HOW TO RECLAIM YOUR PRIVACY. Retrieved from

https://theintercept.com/2015/11/12/edward-snowden-explains-how-to-reclaim-your-privacy/

Mehta, A., Saberi, A., Vazirani, U., & Vazirani, V. (2007). Adwords and generalized online matching. *Journal of the ACM (JACM), 54*(5), 22.

Mihalcik, C. (2016). New York Times, BBC and others inadvertently serve up dangerous ads. Retrieved from https://www.cnet.com/news/new-york-times-bbc-dangerous-ads-ransomware-malvertising/

Mimoso, M. (2015). Ad networks ripe for abuse via malvertising. Retrieved from https://threatpost.com/ad-networks-ripe-for-abuse-via-malvertising/111840/

Patrizio, A. (2016). How Forbes inadvertently proved the anti-malware value of ad blockers. Retrieved from http://www.networkworld.com/article/3021113/security/forbes-malware-ad-blocker-advertisements.html

Safe Internet Use. (2016). Retrieved from https://www.getsafeonline.org/protecting-your-computer/safe-internet-use/

Salusky, W. (2007, Dec 06). Malvertising. Retrieved from https://isc.sans.edu/diary/Malvertising/3727

Schneier, B. (2016, February 23). The Ads Versus Ad Blockers Arms Race. Retrieved from https://www.schneier.com/blog/archives/2016/02/the_ads_vs_ad_b.html

Segura, J. (2015). Real-time Bidding and Malvertising: A case study. Retrieved from https://blog.malwarebytes.org/malvertising-2/2015/04/real-time-bidding-and-malvertising-a-case-study/

Segura, J. (2016). MSN Home Page Drops More Malware Via Malvertising. *MalwareBytes Blog*. Retrieved from https://blog.malwarebytes.org/malvertising-2/2016/01/msn-home-page-drops-more-malware-via-malvertising/

Spam & Phishing. (2016). Retrieved from https://staysafeonline.org/stay-safe-online/keep-a-clean-machine/spam-and-phishing

What Is An Untrustworthy Supply Chain Costing The U.S. Digital Advertising Industry? (2015, February 26, 2016). Retrieved from http://www.iab.com/insights/what-is-an-untrustworthy-supply-chain-costing-the-u-s-digital-advertising-industry/

Zarras, A., Kapravelos, A., Stringhini, G., Holz, T., Kruegel, C., & Vigna, G. (2014). *The Dark Alleys of Madison Avenue: Understanding Malicious Advertisements*. Paper presented at the Proceedings of the 2014 Conference on Internet Measurement Conference, Vancouver, BC, Canada.

Zorabedian, J. (2016). Wired to ad blocker users: pay up for ad-free site or you get nothing. Retrieved from https://nakedsecurity.sophos.com/2016/02/10/wired-to-ad-blocker-users-pay-up-for-ad-free-site-or-you-get-nothing/
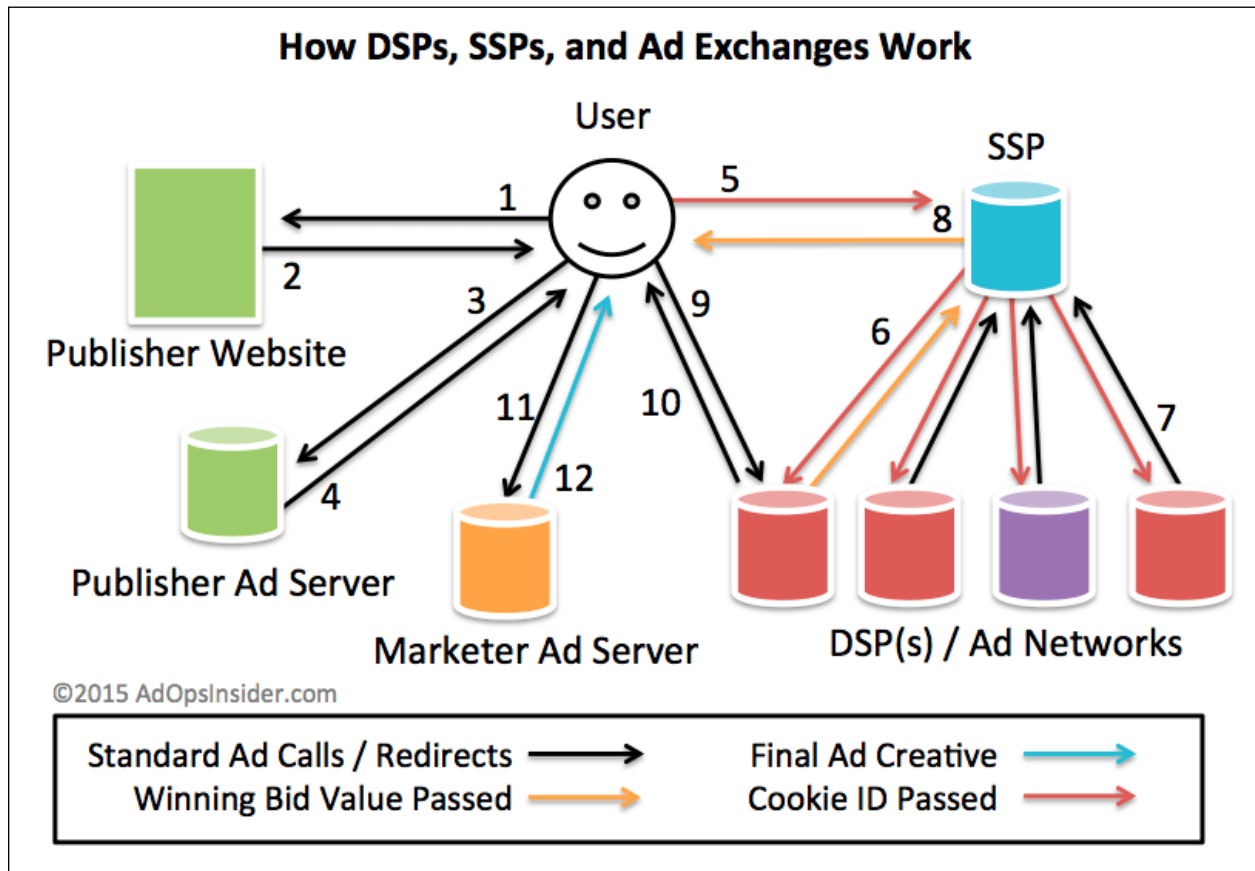
Appendix

Figure 1: How DSPs, SSPs and Ad Exchanges work

Figure 2: Screenshot of Reuters website hacked through an Ad Exchange network

Sentiment Analysis and Opinion Mining: Current State of the Art and Review of Google and Yahoo Search Engines' Privacy Policies

Alan R. Peslak
arp14@psu.edu
Penn State University – Worthington Scranton
Dunmore, PA 18512, USA

Abstract

Sentiment analysis is the review of written or verbal communications to determine some measure of emotion or feeling in the communication. Search engines are one of the most popular sites visited on the Internet generating hundreds of billions of hits per month worldwide. Obviously privacy policies related to these search sites are extremely important. Our study reviews the privacy policies of the two largest US based search engines, Google and Yahoo to determine the overall sentiment of their privacy policies. Significant individual findings and significant differences were found using several sentiment and opinion analysis methods.

Keywords: sentiment analysis, opinion mining, search engines, Google, content analysis, qualitative analysis

1. INTRODUCTION

The field of sentiment analysis and opinion mining is exploding. There is a virtual flood of qualitative data available from a wide variety of sources on the web that can be used to analyze the attitudes behind textual material. Millions of Twitter posts or tweets, millions of Facebook posts and billions of web pages and other documents can be reviewed to determine the opinions behind the words. This analysis can be extremely useful for both researchers and practitioners. Marketing professionals can monitor text communications to determine current attitudes towards their products. Politicians can analyze text communications to determine popularity and feelings toward their candidacy and their stands on issues. Researchers can likewise study text data to find differences, patterns, or trends in a wide variety of text, from policies to presentations, from documents to websites.

This manuscript presents an overall review of the current state of the art in sentiment and opinion analysis. It begins with a review of sentiment analysis including its definition, history, and a review of the literature. Following this is a review of current tool terms and dictionaries that are used in contemporary sentiment and opinion analysis tools. Finally, a detailed example of the use of these tools is presented comparing the sentiment of the privacy policies of two major search engines, Google and Yahoo. A statistical comparison is made of the sentiment results of these two documents and statistical conclusions are made with regard to their sentiment differences.

2. LITERATURE REVIEW

Sentiment analysis has been used extensively in current research. Applications have ranged from education to health care quality to mental health to student performance to customer feedback to politics to product reviews.

One of the most cited and major works dealing with Sentiment Analysis is Sentiment Analysis and Opinion Mining by Bing Liu (2012). In the first chapter he defines the domain. "Sentiment analysis, also called opinion mining, is the field of study that analyzes people's opinions,

sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes." Sentiment Analysis is the review of written or other forms of communication or qualitative data to determine a quantifiable and comparable measure of some form of feeling in the communication or data.

Pang and Lee (2008) suggest that one of the most studied areas of sentiment analysis is sentiment polarity and degree of positivity. A simple concept is to determine whether a particular communication is positive or negative. Eguchi and Lavrenko (2006) note this can be done for summarizing an overall document or retrieving selected sentiment text. "One of the first and still most used method of sentiment analysis is keyword analysis, where a text is reviewed word by word and compared against a dictionary. This dictionary has been previously prepared and will classify each word in its dictionary into a sentiment rating. As an example, good would have a high positive sentiment rating and bad would have a low sentiment rating. But this common analysis has some obvious flaws." Cambria, Schuller, Xia and Havasi (2013). They note two problems with keyword analysis. "Keyword spotting is weak in two areas: it can't reliably recognize affect-negated words, and it relies on surface features. "

Nasukawa and Yi (2003) studied 8 popular sentiment analysis implementations including LIWC and SentiNEt and found some wide variances of polarity between the different methods. This, therefore, suggests that there is not a clear answer when it comes to sentiment ratings and polarity and results should be verified with alternative methods and compared to each other to obtain agreement prior to making definitive conclusions.

Many studies have been performed on privacy policies of Internet sites. Jensen and Potts (2004) examined privacy policies as a decision making rule. Miyazaki and Krishnamurthy (2002) studied the relationship between privacy policies and consumer perception. There is considerable research as well on the inclusion of fair information practices into privacy policies.

There has also been much research on Google and other search engines privacy policies.

Tene (2007) detailed legal issues associated with the Google search site. Piper (2005) warns of the data collection via use of the Google search engine. Zimmer (2008) examined search engine privacy threats.

After a comprehensive Google Scholar search, we could find no instances of sentiment analysis or opinion mining of privacy policies in the literature. With so much information and so much activity, analysis of the major search engines is a fertile area of research. Our review is to analyze privacy statements of the major search engines with regard to sentiment analysis.

3. METHODOLOGY

There are many sources and algorithms for sentiment analysis. But all sentiment analysis include some uncertainty since absolute sentiment analysis is not possible at this time. There are many reasons for this including negation, sarcasm, word combinations, and relative subjectivity. As a result it is useful to use multiple measures to confirm any sentiment analysis findings. This is the approach we have taken. In this way we can both demonstrate and explain different approaches as well as confirm our findings with several sources and algorithms.

There are a variety of tools that can be used for sentiment analysis. We will examine three major tools and show how they measure sentiment, compare results where, and perform a review of two the most popular search engines privacy policies, Google and Yahoo and analyze their sentiment with these tools. In addition, we will discuss an excel lookup function using a significantly larger dictionary than almost all current sentiment analysis engines. We have developed an excel VBA worksheet that analyzes documents using this greatly expanded dictionary and compare these results to traditional sentiment analysis tools. We will also perform statistical analysis t-test to determine if there is significant difference in specific sentiment dimensions in these two privacy policies. This will serve as a review and example of available sentiment analyses and how they can be utilized for qualitative document and communication analysis.

For our analysis, we utilize several online and commercial products as well as several new analyses we have developed using extensive new affective word dictionaries. Specifically we used: RIOT (Recursive Inspection of Text) SCAN software and AYLIEN software as well as two word analyses using DIC-LSA dictionary (Warriner, Kuperman, and Bryssbaert, 2013) and also norms of valence, arousal, and dominance for 13,915 English lemmas dictionary (Bestgen, Y., & Vincze,

N., 2012). With these latter two analyses we are able to calculate significant differences between the two policies. This was because our self-developed algorithm had values for every word in the document and was thus able to allow calculation of means, standard deviations and perform t-tests on the data.

The Google and Yahoo search engine privacy policies (henceforth known as policies or privacy policies) were download from their respective sites and used to perform all analysis. RIOT SCAN is specific downloadable software that you to specify documents in text format and perform detailed content analysis on your documents. The software contains dozens of dictionaries and tables and returns 536 metrics using these dictionaries and other tools. (Boyd, 2014). Though most of these metrics do not measure sentiment, there are several where sentiment analysis is performed.

Three major sentiment calculations are performed by ANEW (Affective Norms for English Words), Harvard General Inquirer, and Lexicoder Sentiment Dictionary. ANEW is popular Sentiment dictionaries and was developed by Bradley and Lang in 1999. It includes three sentiment measures pleasure (or Valence), Arousal, and Dominance. Other dictionaries such as DIC-LSA have adopted this three measure sentiment categories. The categories are Affective Valence (happy to unhappy), Arousal (excited to calm) and Dominance (in-control to not in control). The scale is 1 to 9 with higher numbers indicating favorable affect (happy, excited, in-control). A text is parsed into individual words which are then mapped to a dictionary. The researchers who develop the dictionary performed surveys to determine relative affect score for each word. The ANEW dictionary includes only 1034 words however, and is thus limited in its generalizability.

The Harvard General Inquirer sentiment rating has two separate measures, one for positive words and another for negative words. The current version of the dictionary is extensive now including over 11,000 words (Guerini, Gatti, and Turchi, 2013). The measure calculated is the percentage of words that can be classified as positive or negative words in the entire document.

Young and Soroka (2012) developed their own freely available Lexicoder Sentiment Dictionary in 2012. It includes 4567 positive and negative words and was developed for analysis of news stories related to politics. One of the unique output measures from Lexicoder is a net positive

and negative percentage. It reviews the text for positive and negative words but also for negated positives or negatives and then reverses these to their proper categories. As such, it is a unique addition to the problem of negated words.

A popular sentiment analysis tool is available from the software company AYLIEN. "AYLIEN Text API is a package of Natural Language Processing, Information Retrieval and Machine Learning tools for extracting meaning and insight from textual and visual content with ease." (AYLIEN, 2015) The AYLIEN API analyzes any text and returns a series of sentiment variables to "Detect sentiment of a document in terms of polarity (positive or negative) and subjectivity (subjective or objective)." (AYLIEN, 2016) We used the AYLIEN API plug-in free edition in conjunction with RapidMiner Studio to perform polarity and subjectivity analysis of our policies.

	DIC-LSA Norms
DICLSA_Arousal	cupboard, shelf, fold (low arousal), murderous, violent, coward (high arousal)
DICLSA_Dominan	suffer, loss, victim (low dominance), feat, talent, dedication (high dominance)
DICLSA_Valence	virus, murder, stressful (low/negative valence), enchanting, beauty, dancing (high/positive valence)
DICLSA_Concrete	tomato, spoon, bin (high concreteness), theoretical, imply, vagueness (low concreteness)

Table 1   DIC-LSA Norms

In addition to the external software, we also developed our own VBA enabled Excel spreadsheet in conjunction with two freely available sentiment dictionaries to independently determine sentiment in our policies. This approach also allowed us to determine statistical significance of the differences found between the policies. The two dictionaries used were DIC-LSA (Dictionary Latent Semantic Analysis) and WKB (Warriner, Kuperman, and Brysbaert). The DIC-LSA Norms with example words are presented below. The metrics are all averages of ratings based on the dictionaries. For Concreteness,

higher scores = more concreteness, lower scores = more abstractness. WKB are similar to ANEW.

Metric	Software used	Measures
ANEW All Valence Mean	RIOT	Valence (Positive/Negative)
ANEW All Arousal Mean	RIOT	Arousal
ANEW All Dominance Mean	RIOT	Dominance
Harvard General Inquirer Positive	RIOT	Valence Positive
Harvard General Inquirer Negative	RIOT	Valence Negative
Lexicoder (LSD) Positive Final	RIOT	Valence Positive
Lexicoder (LSD) Positive Final	RIOT	Valence Negative
AYLIEN Polarity	RapidMiner and AYLIEN	Valence
DICLSA Valence	Authors and Dictionary	Valence
DICLSA Arousal	Authors and Dictionary	Arousal
DICLSA Dominance	Authors and Dictionary	Dominance
DICLSA Concreteness	Authors and Dictionary	Concreteness
WKB Valence	Authors and Dictionary	Valence
WKB Arousal	Authors and Dictionary	Arousal
WKB Dominance	Authors and Dictionary	Dominance

Table 2 Metrics Used

Overall, we used twelve separate measures used a variety of software to determine and confirm sentiment analysis of our policies. These twelve measures are summarized in table 2.

4. RESULTS

The results were processed using the software noted and for DICSLA and WKB using SPSS 23 for t-test of independent samples. The numerical results were obtained from analyzing the ratings from the respective metrics of each software product. Each word in the respective document is processed through the dictionary in each dictionary and assigned a scalar variable. These results are either averaged for items such as ANEW_All_Val or shown as percentages as in HARV_Positiv. The scales for RIOT SCAN are shown in table 3.

	Google	Yahoo	Scale
ANEW_All_Val	6.068	6.342	(1-negative to 9-positive)
ANEWAll_Arous	4.539	4.675	(1-calm to 9-excited)
ANEW_All_Dom	5.541	5.573	(1-controlld to 9-in-control)
HARV_Positiv	8.073	7.466	% of positive words
HARV_Negativ	1.790	1.333	% of negative words
LSD_Positive_Final	4.212	4.733	% positive and negated negative words
LSD_Negative_Final	1.088	0.866	% negative and negated positive words

Table 3 RIOT Scan

Recursive Inspection of Text results are presented in table 3. ANEW results show that both Google and Yahoo have positive sentiment scores reflecting generally favorable emotional tone such as happy and pleasant at a 6 on a 1-9 scale. Yahoo has a slightly more pleasant tone. Excitement for both Google and Yahoo are about neutral, neither excited nor calm. Yahoo has marginally more stimulated content. Finally,

Dominance for both show somewhat controlled tone at 5.5 on a scale of 1-9. Yahoo is marginally more dominant.

Harvard results reflect a percentage of positive versus negative words. According to this measure, Both Google and Yahoo have much higher positive words than negative words. The gap is wider in this case for Google resulting in a higher net positive emotional rating for Google.

Finally, the LSD results adjust for the inclusion of negated positive words (e.g. not good) and include them in negative words and vice versa. The results still show that both policies have more positive words than negative and thus are strongly positive in tone. Here though, Yahoo shows a more net positive tone.

Google	Yahoo
polarity : positive,	polarity : positive,
subjectivity : unknown,	subjectivity : unknown,
text : Welcome to the Google Privacy Policy When you use Google services, you trust us with your information. This Privacy Poli...,	text : Welcome to the Yahoo Privacy Center -- take a look around. You'll learn how Yahoo treats your personal information, alon...,
polarity_confidence : 0.98328690807	polarity_confidence : 0.984126984
subjectivity_confidence : 0	subjectivity_confidence : 0

Table 4 AYLIEN results

The results of the RapidMiner with ALYLIEN plugin are presented in table 4. Here we have a black box comparison of polarity (positive or negative, Valence) judgment and a polarity confidence. Both Google and Yahoo are calculated to be positive documents and they both have very high polarity confidence which is the measure of certainty of the polarity determination of positive and negative. Yahoo polarity confidence is slightly higher. Neither policy provides enough information to determine subjectivity levels of the texts (subjective to objective).

Access the DIC-LSA and WKB dictionaries allowed us to perform a word by word analysis of each of our policies. This allowed us to perform statistical analysis of the differences between each policy with regard to the overall metrics calculated. The DIC-LSA Valence results are shown in table 5 and 6.

	GorY	N	Mean	Std. Deviation	Std. Error Mean
V	1.00	1966	5.8669	.53241	.01201
	2.00	902	5.8527	.54557	.01817

Table 5 DIC-LSA Valence (1=Google, 2=Yahoo)

		Levene's Test			
		F	Sig.	t	Sig. (2-tailed)
V	Equal variances assumed	1.382	.240	.658	.511
	Equal variances not assumed			.652	.514

Table 6 DIC-LSA Valence t test for variance

Both Google and Yahoo scored somewhat high in overall valence with scores of nearly 6 on a 9 point scale. The independent samples t-test reveals that the difference between the valences of each is not significant at $p < .05$ or $p < .10$. There is no significant difference in valence between Google and Yahoo privacy policies.

	GorY	N	Mean	Std. Deviation	Std. Error Mean
A	1.00	1966	5.1474	.20949	.00472
	2.00	902	5.1580	.18981	.00632

Table 7 DIC-LSA Arousal (1=Google, 2=Yahoo)

		Levene's Test			
		F	Sig.	t	Sig. (2-tailed)
A	Equal variances assumed	5.174	.023	-1.297	.195
	Equal variances not assumed			-1.345	.179

Table 8 DIC-LSA Arousal t test for variance

Both Google and Yahoo scored somewhat neutral in overall arousal with scores of nearly 5 on a 9 point scale. The independent samples t-test reveals that the difference between the arousal of each is not significant at $p < .05$ or $p < .10$. There is no significant difference in arousal between Google and Yahoo privacy policies.

	GorY	N	Mean	Std. Deviation	Std. Error Mean
D	1.00	1966	5.4118	.28251	.00637
	2.00	902	5.4054	.29376	.00978

Table 9 DIC-LSA Dominance (1=Google, 2=Yahoo)

		Levene's Test			
		F	Sig.	t	Sig. (2-tailed)
D	Equal variances assumed	.001	.982	.549	.583
	Equal variances not assumed			.541	.589

Table 10 DIC-LSA Dominance t test for variance

Both Google and Yahoo scored somewhat positive in overall dominance with scores of nearly 5.5 on a 9 point scale. The independent samples t-test reveals that the difference between the dominance of each is not significant at p < .05 or p < .10. There is no significant difference in dominance between Google and Yahoo privacy policies.

	GorY	N	Mean	Std. Deviation	Std. Error Mean
C	1.00	1966	4.0287	.35063	.00791
	2.00	902	4.0275	.32965	.01098

Table 11 DIC-LSA Concreteness (1=Google, 2=Yahoo)

		Levene's Test			
		F	Sig.	t	Sig. (2-tailed)
C	Equal variances assumed	1.663	.197	.090	.928
	Equal variances not assumed			.092	.927

Table 12 DIC-LSA Concreteness t test for variance

A unique metric in the DIC-LSA dictionary is concreteness. Concreteness is a measure of whether the text is abstract or has definitive or concrete tone. Both Google and Yahoo scored somewhat abstract in overall concreteness with

scores of nearly 5.5 on a 9 point scale. The independent samples t-test reveals that the difference between the concreteness of each is not significant at p < .05 or p < .10. There is no significant difference in concreteness between Google and Yahoo privacy policies.

In our analysis of Google and Yahoo privacy policies using the WKB dictionary, we again had access to their dictionary and performed our own word by word analysis. The statistical analyses that resulted from this comparison are shown in this section.

In order to analyze whether there is a statistically significant difference between the mean Valence of Google's and Yahoo's privacy policies, an independent samples t-test is performed. The results are presented in table 13 and 14. The first item needing to be analyzed is Levene's Test for Equality of Means. Since the significance here is not p < .05, we can conclude that there is not a significance between the variances in the Google and Yahoo data. We therefore need to evaluate the t-test for Equality of Means with the "Equal Variances assumed" row. The t-test significance is p < .012. This result is that the difference between the means for Arousal are significant and Google has a significantly higher Valence sentiment than Yahoo. Google has a more positive sentiment than Google at 5.81 versus 5.68 on a 9 point scale.

	GoogYaho	N	Mean	Std. Deviation	Std. Error Mean
V	1.00	894	5.8071	.83748	.02801
	2.00	442	5.6849	.81957	.03898

Table 13 WKB Valence (1=Google, 2=Yahoo)

		Levene's Test			
		F	Sig.	t	Sig. (2-tailed)
V	Equal variances assumed	.920	.338	2.527	.012
	Equal variances not assumed			2.545	.011

Table 14 WKB Valence t test for variance

	GoogYaho	N	Mean	Std. Deviation	Std. Error Mean
A	1.00	894	3.6199	.75372	.02521
	2.00	442	3.9026	.96875	.04608

Table 15 WKB Arousal (1=Google, 2=Yahoo)

		Levene's Test			
		F	Sig.	t	Sig. (2-tailed)
A	Equal variances assumed	44.163	.000	-5.852	.000
	Equal variances not assumed			-5.383	.000

Table 16 WKB Arousal (1=Google, 2=Yahoo)

In order to analyze whether there is a statistically significant difference between the mean Arousal of Google's and Yahoo's privacy policies, an independent samples t-test is performed. The results are presented in table 1. The first item needing to be analyzed is Levene's Test for Equality of Means. Since the significance here is $p < .001$, we can conclude that there is a significance between the variances in the Google and yahoo data. We therefore need to evaluate the t-test for Equality of Means with the "Equal Variances not assumed" row. The t-test significance is $p < .001$. This result is that the difference between the means for Arousal are significant and there is a significant difference between Yahoo and Google in Arousal sentiment. Both are low in Arousal at 3.90 and 3.62 but Yahoo is significantly more aroused than Google.

	GoogYaho	N	Mean	Std. Deviation	Std. Error Mean
D	1.00	894	5.7700	.73987	.02474
	2.00	442	5.7578	.69715	.03316

Table 17 WKB Dominance (1=Google, 2=Yahoo)

In order to analyze whether there is a statistically significant difference between the mean Dominance of Google's and Yahoo's privacy policies, an independent samples t-test is performed. The results are presented in table 17 and 18. The first item needing to be analyzed is Levene's Test for Equality of Means. Since the significance here is $p < .117$, we cannot conclude

that there is a significance between the variances in the Google and yahoo data. We therefore need to evaluate the t-test for Equality of Means with the "Equal Variances assumed" row. The t-test significance is $p < .772$. This result is that the difference between the means for Dominance are not significant and there is no significant difference between Yahoo and Google in Dominance sentiment. Both Google and Yahoo has a high In Control Sentiment at 5.77 and 5.7578.

		Levene's Test			
		F	Sig.	t	Sig. (2-tailed)
D	Equal variances assumed	2.465	.117	.290	.772
	Equal variances not assumed			.296	.767

Table 18 WKB Dominance t test for variance

5. DISCUSSION

The overall results of our sentiment analysis of the two major search engines privacy policy yielded interesting and mostly consistent results. These results are summarized in table 19 and shown in detail in Appendix A.

Overall, both policies have positive valence or sentiment. The six various analyses are fairly evenly split on which has higher positive sentiment though. The only metric shown to be statistically significant in valence was the WKB Valence which showed Google slightly higher and statistically significant. The overall consensus though appears to be that there is little difference in the positive valence but slightly more in Google. This first metric illustrates the variability that exists among the sentiment tools. The reason for this is that each use different dictionaries. They each have a different number of words in their dictionary and they all have done their own survey to determine sentiment ratings.

The second metric calculated was the Arousal metric. ANEW, DICLSA, and WKB all calculated a level of excitement or arousal for the policies. In two analyses, both policies were found to be neutral in arousal. The WKB results showed slightly less arousal than neutral. In all three, Yahoo showed a higher arousal level and it was statistically significant in WKB. Thus, it can be

said that both policies are neutral to less than neutral arousal and Yahoo is a bit more exciting.

Metric	Measures	Results of Both
ANEW All Valence Mean	Valence (Positive/Negative)	Both somewhat positive
ANEW All Arousal Mean	Arousal	Both Neutral
ANEW All Dominance Mean	Dominance	Both somewhat in-control
Harvard General Inquirer Positive	Valence Positive	Both strongly positive
Harvard General Inquirer Negative	Valence Negative	Both low negative
Lexicoder (LSD) Positive Final	Valence Positive	Both positive
Lexicoder (LSD) Positive Final	Valence Negative	Both low negative
AYLIEN Polarity	Valence	Both Positive
DICLSA Valence	Valence	Both Somewhat positive
DICLSA Arousal	Arousal	Both Neutral
DICLSA Dominance	Dominance	Both somewhat in-control
DICLSA Concreteness	Concreteness	Both somewhat abstract
WKB Valence	Valence	Both somewhat positive
WKB Arousal	Arousal	Both somewhat less aroused
WKB Dominance	Dominance	Both somewhat in-control

Table 19 Summary of Results

ANEW, DICLSA and WKB also provided a measure of Dominance or feeling of being in control. Not surprisingly, both privacy policies which explicitly deal with control showed metrics above neutral and both showed "somewhat in-control". The three analyses are split between which had the higher control and none were statistically significant. We therefore conclude that there was no difference in level of dominance in the policies.

The final metric measured by DICLSA was concreteness. Surprisingly, both policies were less than neutral and both were somewhat abstract, not well defined or concrete. There was no statistical difference between the two policies based on concreteness. One possible reason the policies are abstract to leave room for the companies to have legal flexibility.

6. CONCLUSION

Overall this study has contributed to the literature in three ways, first it defines, presents and demonstrates six different methods of sentiment analysis. Researchers and practitioners can use this manuscript as a source, primer and guide for developing their own sentiment analysis of any communication. Second, the study illustrates the inexact but relatively consistent results that are generated by several sentiment analysis tools and dictionaries. Researchers and practitioners can reliably use any of the tools and obtain similar results regardless of the tools used. Note that there is some small variation that will be experienced. Finally, the study analyzes the privacy policies and sentiment and tone of the two largest search engines. The results show little differences in any of the sentiment measures between Google and Yahoo. Both are somewhat positive in sentiment, neutral in arousal, somewhat in control in dominance, and somewhat abstract documents. Researchers can use these findings to compare to other search engines policies or other privacy policies for other type sites to compare and contrast their sentiment characteristics. Search engines companies can use these findings to improve their overall sentiment if they choose. Potential changes in privacy policies for companies could be to make privacy policies, happier, less controlled, provide a change in arousal, and be more concrete.

7. REFERENCES

AYLIEN (2015).   *Text Analysis API.* http://aylien.com/text-api

AYLIEN (2016) *Sentiment Analysis* http://aylien.com/sentiment-analysis/

Bestgen, Y., & Vincze, N. (2012). Checking and bootstrapping lexical norms by means of word similarity indexes. *Behavior Research Methods*, 44(4), 998-1006. (DIC-LSA)

Boyd, R. L. (2013). RIOT Scan: *Recursive Inspection of Text Scanner (Version .0.11)* [Software]. Available from http://riot.ryanb.cc

Bradley, M. M., & Lang, P. J. (1999). Affective Norms for English Words (ANEW): Stimuli, Instruction Manual and Affective Ratings. Technical report C-1, Gainesville, FL. The Center for Research in Psychophysiology, University of Florida.

Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, (2), 15-21.

Guerini, M., Gatti, L., & Turchi, M. (2013). Sentiment analysis: How to derive prior polarities from SentiWordNet. *arXiv preprint arXiv:1309.5843*.

Jensen, C., & Potts, C. (2004, April). Privacy policies as decision-making tools: an evaluation of online privacy notices. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 471-478). ACM.

K. Eguchi and V. Lavrenko, "Sentiment retrieval using generative models,"in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 345–354, 2006

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167.

Miyazaki, A. D., & Krishnamurthy, S. (2002). Internet seals of approval: Effects on online privacy policies and consumer perceptions. *Journal of Consumer Affairs*, 36(1), 28-49.

Nasukawa, T., & Yi, J. (2003, October). Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of the 2nd international conference on Knowledge capture* (pp. 70-77). ACM.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2), 1-135.

Piper, P. S. (2005). Google and privacy. *Internet reference services quarterly*, 10(3-4), 195-203.

Tene, O. (2007). What google knows: Privacy and internet search engines. *Utah Law Review, Forthcoming*.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior research methods*, 45(4), 1191-1207.

Young, L. and Soroka, S. (2012). Affective news: The automated coding of sentiment in political texts. Political Communication, 29(4), 205-231. LSD

Zimmer, M. (2008). The externalities of search 2.0: The emerging privacy threats when the drive for the perfect search engine meets Web 2.0. *First Monday*, 13(3).

Editor's Note:

Appendix A

Metric	Measures	Results of Both	Higher rated	Significant
ANEW All Valence Mean	Valence (Positive/Negative)	Both somewhat positive	Yahoo	NA
ANEW All Arousal Mean	Arousal	Both Neutral	Yahoo	NA
ANEW All Dominance Mean	Dominance	Both somewhat in-control	Yahoo	NA
Harvard General Inquirer Positive	Valence Positive	Both strongly positive	Google	NA
Harvard General Inquirer Negative	Valence Negative	Both low negative	Yahoo(less negative)	NA
Lexicoder (LSD) Positive Final	Valence Positive	Both positive	Yahoo	NA
Lexicoder (LSD) Positive Final	Valence Negative	Both low negative	Yahoo (low negative)	NA
AYLIEN Polarity	Valence	Both Positive	Yahoo	NA
DICLSA Valence	Valence	Both Somewhat positive	Google	NO
DICLSA Arousal	Arousal	Both Neutral	Yahoo	NO
DICLSA Dominance	Dominance	Both somewhat in-control	Google	NO
DICLSA Concreteness	Concreteness	Both somewhat abstract	Google	NO
WKB Valence	Valence	Both somewhat positive	Google	YES
WKB Arousal	Arousal	Both somewhat less aroused	Yahoo	YES
WKB Dominance	Dominance	Both somewhat in-control	Google	NO