Journal of Information Systems Applied Research

Volume 13, Issue 3
November 2020
ISSN: 1946-1836

In this issue:

iscap
Information Systems & Computing Academic Professionals

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 three issues a year. 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 EDSIG who perform the editorial and review processes for JISAR.

Journal of
Information Systems Applied Research

Geographic Information System and Gerrymandering

Peter Y. Wu
wu@rmu.edu
Department of Computer and Information System

Justin P. DePlato
deplato@rmu.edu
Department of Political Science

Adam B. Combs
combs@rmu.edu
Department of Mathematics

Robert Morris University
Pittsburgh, PA 15108, USA

Abstract

Gerrymandering is the practice of manipulating voting district boundaries to gain political advantage in democratic voting. The geographic information system (GIS) becomes a versatile tool for that. This paper describes how to use the GIS for gerrymandering, in the practice of both cracking - diluting the opponents voting into many districts, and packing - concentrating the opponent's voters into fewer districts. The use of GIS makes extreme gerrymandering relatively easy to do. Even when we understand it to be bad for democracy since it facilitates for the politician to choose his/her voters, gerrymandering is generally allowed by law. Restricting the practice of gerrymandering turns out to be a legally challenging proposition. We discuss some approaches to legislation against gerrymandering. Believing that the GIS can be part of the solution, we call for GIS researchers to work with legal professionals to formulate regulations to contest and disallow gerrymandering.

Keywords: Gerrymandering, GIS, Geographic Information System.

1. INTRODUCTION

Gerrymandering is the practice of manipulating voting district boundaries to gain political advantage in democratic voting (Griffith 1907). Re-districting is necessary for demographic changes such as birth, death, population migration as well as immigration. By law, a census is done every ten years, and the government has the duty and the right to re-districting. The state law in some states may have the stipulation that the adjustment of voting district boundaries is to preserve the democratic election of government by voting. Nevertheless, it is legal for the political party in power to gain political advantage by gerrymandering, since the law was never written specifically to identify it and disallow it.

Partisan gerrymandering was not a serious problem until the past decade. Generally, it was not an easy task to do. However, the use of geographic information system (GIS) along with the availability of data has made it quite practicable (Reitsma 2013). Some attempted to automate the process (Li, Wang & Wang 2007; Yamada 2009; Siegel-Hawley 2013). Quite a few visionary researchers sought to identify it and disallow it (Niemi, Grofman, Carlucci & Hofeller 1990; Flint 2003; Chou & Li 2006; Ricca, Scozzari & Simeone 2008). If partisan gerrymandering is identified, it may be contested in court and legally disallowed. Many are then calling for research in this area (Forest 2018; Crane & Grove 2018; Grofman & Cervas 2018). Following the past effort, this paper describes how the GIS has become the tool for gerrymandering, and suggest that it may become part of the solution with further research.

Section 2 will present a brief history of the term gerrymandering. It was widely perceived as bad for democracy but it has always been legal. Section 3 will summarily explain the two fundamental strategies of gerrymandering: how to do re-districting to gain political advantage. A few simple figures help to explain that. While there is no existing algorithm to automate gerrymandering, the GIS becomes a viable tool to make it easy. Section 4 goes on to describe how to leverage the GIS interactive functionalities, visualization on the map and spatial data analysis to do gerrymandering. Section 5 begins the discussion of how we may prevent the practice of gerrymandering, suggesting various approaches. Some of these are primarily socio-political, but some inevitably involves geographical and social data analysis. Section 6 presents the summary and a statement of conclusion.

2. BRIEF HISTORY

Elbridge Gerry (1744-1814) was a politician among the founding fathers of the United States of America; a portrait of him is below in Figure 1.

In 1812, the Massachusetts state governor Elbridge Gerry (1744-1814) signed a bill that created a voting district in the shape of a salamander, intended to gain political advantage. The map of the proposed voting district is illustrated in Figure 2. Approved by the state legislature, the bill then also coined the term "gerrymandering" to refer to the practice of manipulating voting district boundaries to gain political advantage (Griffith 1907).

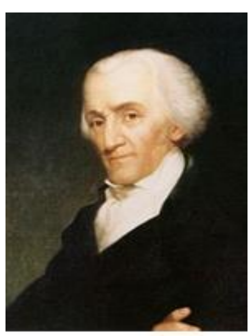


Fig.1 Elbridge Gerry

Periodic re-districting is necessary to allow effective administration of voting by drawing the voting districts according to population distribution. By the U.S. constitution, the federal government cannot dictate how the states may define the voting districts. Every state government sets up its own policy. The political party holding majority in the state government therefore has the privilege to re-draw the voting districts, possibly manipulating that for political gain. Including the US census every 10 years, there is always updated demographic information about the population to justify re-drawing the voting districts. The intention for doing so is difficult to contest.

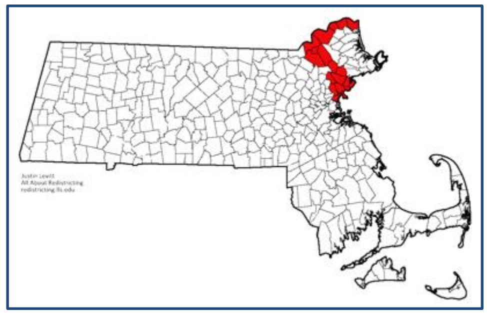


Fig.2 Voting District by Governor Gerry 1812.

Gerrymandering is practically legal. However, it has not been a major issue until more recently, in the past decade. Voting districts drawn for political advantage now begin to show up in evidently very strange shapes and much more often. We believe the common use of GIS today together with the ease of access to data has made it simple to achieve. In the next section, we will discuss the strategies of gerrymandering, and how the GIS makes it much easier.

3. GERRYMANDERING IN PRACTICE

How does the GIS make gerrymandering easy? Let us first examine how to do gerrymandering. Fundamentally, there are two basic strategies: cracking and packing. The choice depends on whether or not the political party has the majority of the votes. Simple illustrations in figures 3, 4 and 5 will explain the ideas quite well.

Suppose the two political parties are A and B. Party A has the majority, 55% of the votes, while Party B has 45%, being the minority. Figure 3 illustrates the hypothetical distribution of the population in a square sample piece of land, and it depicts a simplistic way to form 5 voting districts in five vertical strips. Party A wins 3 districts two of which having 100% of the votes and one district by 75% of the votes. Party B wins 2 districts by 100% of the votes and loses 1 district with only 25% of the votes in one district. Party A has the majority while Party B still has a substantial minority.

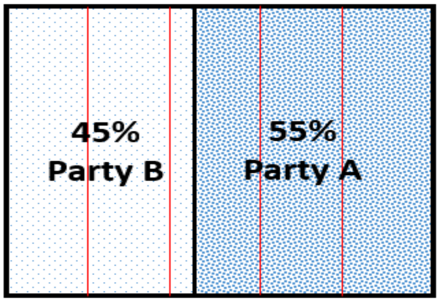


Fig.3 Majority wins 3; Minority 2.

Cracking

Cracking is the approach to dilute the votes of the opposing party to suppress them from winning in any voting district. Suppose Party A is in power, and re-draws the voting districts into 5 horizontal strips, as illustrated in Figure 4. In each of the 5 districts, Party A has the 55% majority and Party B has the 45% minority. Hence Party A wins all 5 districts and Party B loses in all 5 and does not even have a minority say now. The re-districting strategy has distributed the voting power of Party B and suppressed them from winning any district. Cracking is the approach when the party has the majority.

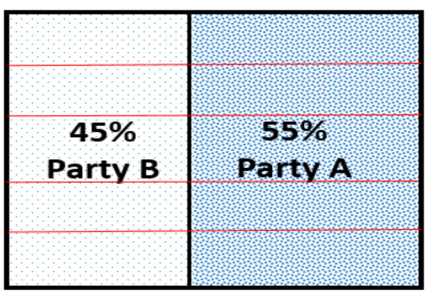


Fig.4 Cracking to Eliminate the Minority.

Packing

Packing is the approach to concentrate the votes of the opposing party in one or few districts to reduce their votes in other districts. Suppose Party B is in power but realizes that they have overall only 45% of the popular votes. In order to gain political advantage, the voting districts are re-drawn, illustrated in Figure 5. One voting district is a vertical strip to the right, with 100% of the votes for Party A. Party A wins the district. But the rest of the area is divided into 4 horizontal strips for the 4 districts. Now in each of these 4 districts, Party B wins by the ratio of 45-to-35, winning in all 4 districts. The result of gerrymandering is that the minority Party B wins 4 districts and the majority Party A wins only one. Packing is the approach when the party has the minority, packing the majority party in one or few districts, reducing their voting power in the rest.

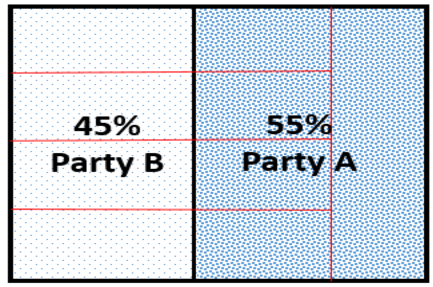


Fig.5 Packing to Limit the Majority.

4. GIS FOR GERRYMANDERING

The simplistic population distribution assumed in our hypothetical map makes it easy for us to explain and illustrate the two fundamental strategies of gerrymandering. In a real situation, it may not be so easy to form the voting districts to achieve cracking or packing. Theoretically, no algorithm exists to exhaustively search for all feasible solutions in gerrymandering.

A better approach is to use the GIS for interactive decision support. We need to first gather the data about where the voters are and which side they are likely to vote for. Such a map presented by the GIS will serve as a visual guide to see where the voters are located. The process is known in the GIS functionality as geocoding (Wu & Rathswohl 2010).

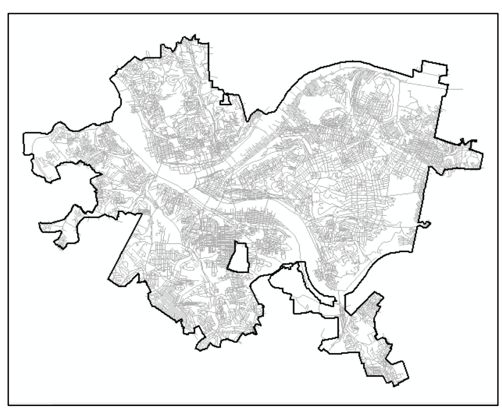


Fig.6 A City Street Map

To illustrate that in a simple way, suppose we have done the foot work of collecting the resident addresses of our political supporters who are to vote for us in a certain city. Figure 6 below shows a street map of the city.

Depending on the level of granularity desired for our map, it may be polygon geocoding just to identify the number of voters on each side within each area unit, or we may apply linear geocoding if we want to identify the point location of each voter (Goldberg 2016). For our illustration, we applied linear geocoding: from the collected addresses, the GIS produces a point map showing where each voter is located by the address. Figure 7 below shows the point map produced by geocoding superimposed on the street map.

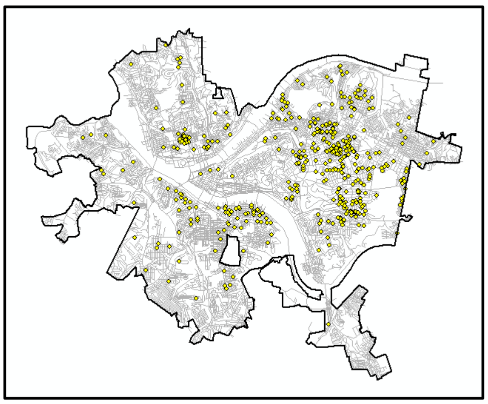


Fig.7 Geocoded Address Locations

Once we have the map to visualize where the voters are, we can use that as our guide to draw the desirable voting district boundaries. Suppose we want to make sure that one substantial group of our supporters in the north east will win in one district. We then draw on the map our desired district, as shown in Figure 8 below.

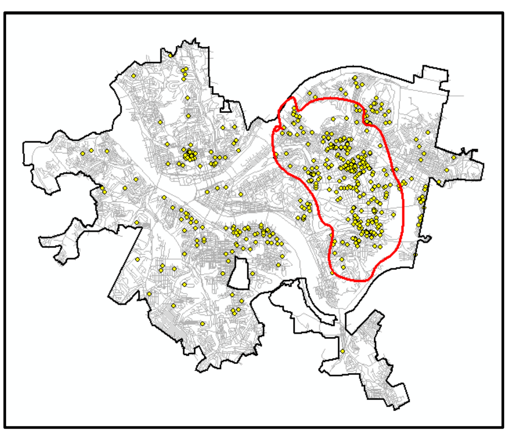


Fig.8 Drawing a voting district

For every district drawn, the functionality known as spatial join of the GIS allows us to immediately calculate the number of our supporters included there, and we can therefore project how likely we may win the voting district.

Suppose we recognize that our supporters do not constitute the majority and therefore need to at least win 3 districts. We can try drawing districts in various shapes, evaluating in each case, until we find the ones we desire. This trial and test approach guided by the visualized map becomes a very practicable way to obtain a robust solution for gerrymandering. Figure 9 shows our desired result of three districts, practicing extreme gerrymandering.

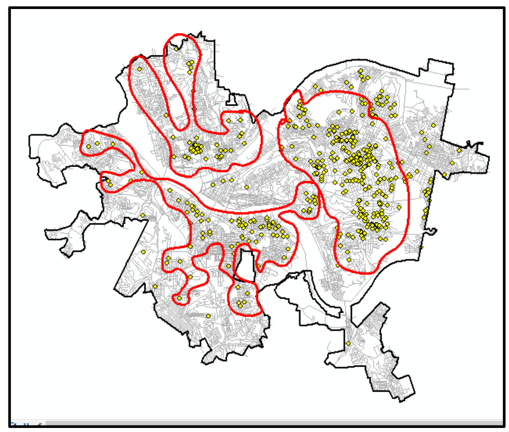


Fig.9 Gerrymandering result in 3 districts

Past attempts to fully automate the process using GIS were not successful (Li, Wang & Wang 2007; Yamada 2009; Siegel-Hawley 2013). There is significant difficulty in traversing or iterating the varieties of options in forming geometric shapes. Whereas using the GIS as an interactive tool for gerrymandering in the computer aided process has been much more promising. In the past decade, we have seen a rising number of cases of extreme gerrymandering (Forest 2018; Crane & Grove 2018).

5. TO PREVENT GERRYMANDERING

Political re-districting is necessary to facilitate for democratic voting when there are changes in the demographics of the voting population. In the past decade, however, gerrymandering becomes a way for politicians in power exercising their rights to deeply entrench themselves with political advantage. It becomes a difficult legal issue how to contest a re-districting map as gerrymandering. There appears to be no easy solution. In this section, we will discuss the issues about how to prevent partisan gerrymandering.

The Electoral College by the founding fathers was originally meant for proportional representation in government. 1824 marked the significant shift to the Winner-Take-All rule in having districts for presidential as well as local elections (Mccarthy 2012). To outline our suggested approaches, this section will discuss the politics of the Winner-Take-All rule and use of non-partisan commission for political re-districting first, followed by the technical issues defining gerrymandering in terms of geometry, and the possible restriction in the use of voter information in re-districting. Finally, the state laws frame the regulation about political re-districting. We think that GIS researchers should work with legal professionals about this, and hopefully, GIS may be part of the solution.

The Winner-Take-All Rule

Gerrymandering is a possible tactic because of the winner-take-all rule. The rule lets the majority winner of a voting district to claim the entire electoral count of the district. Without that rule, gerrymandering will not matter since re-districting will not affect the total count of popular votes. The rule, however, is originally designed to allow a minority population to still have a voice in a democracy when there may be at least some districts within which the minority population becomes a majority. The Voters Rights Act of 1965 requires certain states to ensure minority representation, with at least one district formed based on race (US Department of Justice 1965). Ironically, that is gerrymandering in practice. In US presidential election, some states begin to consider dropping the winner-take-all rule to count only the total popular votes. On the other hand, swing states may then lose their relevance to the candidates if the Electoral College is designated to be proportional to the popular votes. That is for each state to consider. In similar ways, local governments may consider whether or not the winner-take-all rule should be adopted in their specific situations. In either case, political districting needs to preserve a channel for the minority.

Democracy ought to be based on government by proportional representation. However, given that the Winner-Take-All Rule cannot be abolished, we may then seek to revise it for appropriate adoption. It is worthwhile to note that currently in US presidential election, Maine and Nebraska both implement a kind of hybrid combining state wide and district vote counts (Mccarthy 2012).

Non-Partisan Commission

To avoid politicians in power exploiting the opportunity of re-districting by gerrymandering, often suggested is the solution to have a non-partisan commission in charge of re-districting so that there would be no intention to gain political advantage for either side. The idea is simple but the problem is the same. The political hot potato becomes: who should be in that commission? The political problem is only re-casted in a different venue. The Non-Partisan Commission approach is practiced in Florida, but it still has to deal with the requirements of the Voters Rights Act, including the creation of districts for minority representation. The current apportionment rule requires a population of roughly 750,000 per single member district, the practice would be even more difficult when population dispersion is not so ideally congruent, but more clustered than spread out, as in the case of Pennsylvania.

It is unlikely the approach will further involve research in using and understanding GIS, we will not go on with the discussion in this paper.

Recognizing Geometry in Gerrymandering

Founded strongly in theoretical computer science, the field of computational geometry has spawned many algorithms for programming to process geometry represented in digital data (Forrest 1971; Preparata & Shamos 1988). Much of GIS functionality has been built on the results of the research work. In the introduction of this paper, we noted that earlier attempts to automate the process of gerrymandering were not successful. Realizing that extreme gerrymandering is bad for democracy, many researchers then sought to develop programming algorithms to identify partisan gerrymandering so that it might be objectively disallowed. This paper also cited as references some of the research papers. Viewers of a district map may identify weird geometric shapes as evidence of politically motivated re-districting; such evidence is hardly objective proof, much less so in the court of law. Despite many creative esoteric ideas in research, a legal definition for gerrymandering in geometric and demographic terms remains an open question. In a more recent paper, two mathematicians Alexeev and Mixon (2018) summed up in much more definitive terms about how theoretically inconclusive we can expect the approach to be. The paper was titled "An Impossibility Theorem for Gerrymandering."

Perhaps the definitive algorithmic solution is not ready. We however would note that application of artificial intelligence with machine learning to recognize gerrymandering has not yet been much explored.

Restrict the Use of Information

Gerrymandering requires the information about location of the voters as well as their voting inclination. How the information is used in re-districting may expose the intention to gain political advantage through gerrymandering. Legislature may therefore require the appropriate justification for re-districting to indicate that it preserves or promotes democracy. The exact details of such regulation however can become very tricky to articulate, particularly when we may also note the often significant co-relation between other demographic factors such as poverty and wealth, education level, racial and ethnic origin with voting inclination. The regulations will inevitably involve the geographical and analytical issues of population data. We believe this can be a more promising approach for our research so that GIS may be part of the solution to prevent gerrymandering.

State Laws

By U.S. constitution, the federal government will not interfere with how each state may govern the districting of voting population. Two recent Supreme Court cases in June (Maryland and North Carolina) affirmed that interpretation. It is therefore up to each state government to set up the policies for political re-districting. Gerrymandering has been legal since the state laws in general were not specifically written to identify and disallow it. Some states, such as Pennsylvania, may have law stipulating that political re-districting should promote democracy. The recent case of League of Women Voters v. Commonwealth of Pennsylvania (Grofman and Cervas 2018) may shed some light on the issues.

Nevertheless, it is time for GIS researchers to work with legal professionals on the topic, for a better democracy in the future.

6. SUMMARY AND CONCLUSION

Gerrymandering is the practice to manipulating voting district boundaries to gain political advantage in democratic voting. We presented a brief history of the term, and discussed two common approaches in gerrymandering: cracking and packing. Cracking is the approach to dilute the opponent's voting power by distributing the voters into more districts so that the opponent will not win any of the districts. Packing is the approach to concentrate the opponent's voting power into fewer districts so that opponent will win only those districts. Provided with the information where the voters are, the GIS readily presents the map to visually guide our search effort in re-districting. The GIS analytic functionalities can conveniently support trial and test each potential re-districting solution for extreme gerrymandering. To preserve and promote democracy, gerrymandering should be identified and disallowed. But it is quite a challenge to legally define it. Gerrymandering is possible because of the winner-take-all rule in counting votes. The winner-take-all rule is meant to promote democracy by preserving the voice of minority groups. We will have to take that into account. To disallow gerrymandering, the state government must now heed the work of legal professionals working with GIS researchers to identify and disallow gerrymandering.

7. REFERENCES

Alexeev, B. and Mixon, D.G. (2018). An Impossibility Theorem for Gerrymandering. *The American Mathematical Monthly* 125(10), 878-884.
[https://doi.org/10.1080/00029890.2018.1517571]

Chou, C.I. and Li, S.P. (2006). Taming the Gerrymander - Statistical Physics Approach to Political Districting Problem. *Physica A: Statistical Mechanics and Its Applications*, Elsevier.
[https://doi.org/10.1016/j.physa.2006.01.082]

Crane, N.J. and Grove, K. (2018). An Active Role For Political Geography In Our Current Conjuncture. *Geography Compass* 12(11). Wiley Online Library.
[https://doi.org/10.1111/gec3.12410]

Flint, C. (2003). Political Geography: Context and Agency in A Multiscalar Framework. *Progress in Human Geography* 27(5), 627-636.
[https://doi.org/10.1191/0309132503ph453pr]

Forest, B. (2018). Electoral Geography: From Mapping Votes To Representing Power. *Geography Compass* 12(1). Wiley Online Library.
[https://doi.org/10.1111/gec3.12352]

Forrest, A.R. (1971). Computational Geometry. *Proceedings of Royal Society London* 321(4), 187-195.

Fulmer, S. (2019). Pennsylvania Population. Published online by World Population Review, Walnut, CA.
[http://worldpopulationreview.com/states/pennsylvania]

Goldberg, D.W. (2016). Geocoding. *The International Encyclopedia of Geography*, D. Richardson (eds. et al). John Wiley & Sons.
[https://doi.org/10.1002/9781118786352.wbieg1051]

Griffith, E.C. (1907). *The Rise And Development Of The Gerrymander*. Scott, Foresman and Company, Chicago, IL.

Grofman, B. and Cervas, J.R. (2018). Can State Courts Cure Partisan Gerrymandering: Lessons from League of Women Voters V. Commonwealth of Pennsylvania. *Election Law Journal: Rules, Politics, and Policy* 17(4), 264-285. Mary Ann Liebert, Inc., publishers.
[https://doi.org/10.1089/elj.2018.0496]

Li, Z., Wang, R-S. and Wang, Y. (2007). A Quadratic Programming Model For Political Districting Problem. *The First International Symposium on Optimization and Systems Biology* (OSB'07), Beijing, China.

Mccarthy, D. (2012). How the Electoral College Became Winner-Take-All. Published online by FairVote, Takoma Park, MD.
[https://www.fairvote.org/how-the-electoral-college-became-winner-take-all]

Niemi, R.G., Grofman, B., Carlucci, C. and Hofeller, T. (1990). Measuring Compactness and The Role of a Compactness Standard in a Test For Partisan and Racial Gerrymandering. *The Journal of Politics* 52(4), 1155-1181. University of Chicago Press.
[https://doi.org/10.2307/2131686]

Preparata, F.P. and Shamos, M.I. (1985). *Computational Geometry – An Introduction*. Springer-Verlag. ISBN 0-387-96131-3.

Reitsma, F. (2013). Revisting The 'Is GIScience a science?' debate (or quite possibly scientific gerrymandering). *International Journal of Geographical Information Science* 27(2), 211-221.
[https://doi.org/10.1080/13658816.2012.674529]

Ricca, F., Scozzari, A. and Simeone, B. (2008). Weighted Voronoi Region Algorithms For Political Districting. *Mathematical and Computer Modeling* 48(9-10), 1468-1477.
[https://doi.org/10.1016/j.mcm.2008.05.041]

Siegel-Hawley, G. (2013). Educational Gerrymandering? Race and Attendance Boundaries in a Demographically Changing Suburb. *Harvard Educational Review* 83(4), 580-612.
[https://doi.org/10.17763/haer.83.4.k385375245677131]

United States Department of Justice (1965). History of federal Voting Rights Laws: The Voting Rights Act of 1965.
[http://justice.gov/crt/history-federal-voting-rights-laws]

Wu, P.Y. and Rathswohl, E.J. (2010). Address Matching: An Expert System and Decision Support Application for GIS. *Proceedings of Information Systems Education Conference* (ISECON 2010). ISSN: 1542-7382, #1339, Nashville, TN.
[http://proc.isecon.org/2010/pdf/1399.pdf]

Yamada, T. (2009). A Mini-Max Spanning Forest Approach To The Political Districting Problem. *International Journal of System Science* 40(5), 471-477.
[https://doi.org/10.1080/00207720802645246]

Knowledge Management System Development: Handling Evolution from Explicit to Tacit and Social

William H. Money
wmoney@citadel.edu
Baker School of Business
The Citadel
Charleston, SC 29409, USA

Lionel Mew
lmew@richmond.edu
Information Systems
University of Richmond
Richmond, VA 23173, USA

Abstract

This paper investigates evolution of Knowledge Management (KM) systems development approaches. KM has been an important topic for over forty years. Period KM systems development has been based on definitions of knowledge that have evolved, with systems developed using tools and approaches characteristic of the time. The first KMS were gateways to static information supporting integrative business processes. As KM evolved, knowledge became seen as dynamic, moving within the enterprise due to organizational processes. Current thought focuses on tacit knowledge, which is hard to explicitly transfer, making KM a social process. Analytics will increasingly play a factor in new KM systems. The question is whether development approaches have kept up with evolving definitions of knowledge. This paper identifies approaches implementing new KM systems, draws on the literature to examine how they are used, and discusses whether they reflect knowledge's changing nature. The methodologies used to develop KM systems are examined with the goal of providing insight into approaches that work.

Keywords: Knowledge Management, System Development, Paradigm Shift

1. INTRODUCTION

This paper discusses how KMS development methodologies have changed, but not necessarily kept up with changing (expanding) definitions of knowledge. As our concept of knowledge has broadened from a static content model to a more dynamic model followed by social content paradigms (Tzortzaki & Mihiotis, 2014), a review of the research suggests that there has been some change in the way KMS are constructed. However, a question is whether KM system development methodologies have evolved purposefully as our understanding of knowledge has expanded, or by responding simply to development advances and not in response to knowledge paradigm shifts.

We do not delve deeply into the conceptualization of knowledge and its history in this paper. It is well recognized, since the 17th century philosophers (led by Descartes' work) have approached knowledge as involving human acceptance of facts and an understanding that something is not in doubt, or has some large degree of certainty. If the certainty is extremely high, it is then knowledge, accredited as certain, and not doubted (Newman, 2008).

Researchers also broadly agree that information and knowledge are closely linked, with information being accepted as categorized, or meaningfully arranged data. Knowledge, according to Denning (1998), is an extension of this data – the understanding of information, for after information becomes certain, we have knowledge. Knowledge exists for humans where something (a question or understanding) is believed to be true or certain (Denning, 1998).

Information systems are characterized that are designed and developed to manage organizational knowledge as Knowledge Management Systems (KMS). These KMS may functionally aid organizational learning, ingest and store organizational knowledge, and make it accessible as required for recall and application (Damodaran & Olphert, 2000). The content of a KMS may include corporate history, project and personnel experience and expertise, and the knowledge that promoted the success of the business.

In the late 1980's, understanding of knowledge was advanced with the conceptualization of a knowledge hierarchy by Ackoff (1989) where data led to information, knowledge, and finally wisdom in a hierarchy depicted as a DIKW pyramid. The model has been discussed, with definitions argued and examined in numerous papers (for example see (Frické, 2009; Hey, 2004; Sharma, 2008; Tuomi, 1999). However, until the 2009 time period, the discussion and papers did not begin to deeply analyze or propose succinct step-like or phases for processes that would be used to describe how a KMS could be constructed, and the systematic function a KMS would perform.

Jennex (2009) dissected and discussed the pyramid to clarify the definitions used and offered insight and design inputs and outputs to advance KMS construction. He identified the processing sequence differences (upward form Ackoff's pyramid design) and downward (from Tuomi's) available; postulated and gave clear use case examples showing that a revised knowledge pyramid could have bi-directional flows; validation mechanisms (for KM strategy); incorporating social networks (data creation and transport); filtering; communication; collaboration and work processing elements. Jennex points out how KMS are functionally more than just knowledge storage and retrieval technologies.

Although metrics tend to vary, it is widely accepted that KMS performance measurement revolves around how effective the systems are in terms of meeting organizational goals, while enhancing individual performance and satisfaction. The long history of knowledge studies has a practical significance for organizational work because it greatly impacts the ability of people and organizations to understand and act effectively. Organizations must survive in competitive environments, and assemble knowledge to support organizational processes, promote effective functioning and provide valuable assets for sale or exchange. As both knowledge and competitors improve, KMS practices and support must also improve. They must recognize the developments in technology and people-centric areas to continue success (Wiig, 2000).

The approaches and activities for capturing and managing knowledge have been undertaken frequently as practical projects targeted at providing direct support for organizational objectives with a clear understanding of underlying organizational processes that are implemented with or directly supported by the relevant KMS. These KM projects are not an attempt to construct organization wide KMS for they focus KM efforts directly on organizational needs and capabilities by constructing so-called adaptive, contextual, comprehensive, and people-centric types of environments that focus on knowledge-related concerns (Wiig, 2000).

These approaches have resulted in initiatives to increase knowledge sharing among individuals by building instructional and learning programs and knowledge distribution capabilities; manage knowledge through capturing, manipulating, and locating knowledge; and on knowledge utilization by building and exploiting information management to improve enterprise economic value. Finally, some have developed into more widely used tools where information and knowledge is more broadly utilized and exploited as a central resource. These KMS function as created environments which focus constant, widespread organizational attention on ensuring competitive information is available to sustain long-term success and viability (Wiig, 2000). Efforts have thus been directed at turning tacit knowledge into explicit knowledge with the straightforward capture and sharing methods.

As the definition and use of knowledge has expanded from static to dynamic to social, research shows that the tools used to develop KMS have changed, but not necessarily in concert with our expansion of the new definitions associated with the knowledge pyramid, or a design based on an understanding of the many different function, and objects of a KMS. Little research is available discussing how knowledge is used in organizations, how knowledge supports organizational goals, and that the use cases for knowledge that would be required to develop robust KMS are not readily available or well validated. It is suggested that a deeper understanding of how people interact with knowledge must be gained, and that methodologies for developing KMS be updated to reflect that understanding.

2. WHAT IS KNOWLEDGE?

The definitions and theories of knowledge are continuously evolving. These changes in the understanding of knowledge are in turn modifying and adding to what is incorporated into the KMS, and how the KMS must be constructed and managed as they grow. For the previous 20 years or more, Davenport, De Long, & Beers' 1998 description of how experience, context, interpretation, and reflection become knowledge when united with information has been a dominant theme. The resulting KMS products have been important for both decisions and actions. Unfortunately, differentiating between the information and knowledge is both difficult and problematic in practice, but it is apparent that human input is important as data becomes information and information then becomes knowledge (Davenport, De Long, & Beers (1998).

How Content affects the KMS

KMS based on this conceptualization of knowledge are constructed and used for a basic purpose that has become increasing complex. They are designed to collect, hold, and when called upon (by humans or decision tools) – deliver knowledge to users. These focused knowledge management functions are valuable because they leverage the inputs (data), support analysis, may also contain experience, and individual knowledge inside and outside of an organization (Ruggles, 1998). As the types of data included in the systems has expanded to include Big Data with unstructured text, sensor outputs, and social media. In 2013, Jennex & Bartczak conceptualized a revised knowledge pyramid to describe the knowledge that may be included in KMS of today. These authors postulate that the actions of modern systems incorporate learning, filtering, and transformation processes to generate a significant difference between the KM knowledge pyramid and the earlier general knowledge pyramid (Jennex & Bartczak, 2013).

The 2013 Jennex a& Bartczak pyramid describes the actions through which a KM delivers actionable intelligence and identified filters, functional processes, and technologies as being integral to the delivery process. However, the authors note that KM, as comprehended and depicted through a knowledge pyramid, does not incorporate Big Data, analytics, and the Internet of Things. In a more recent 2017 paper, Jennex recognizes how understanding of the data has expanded and utilized this KM conceptualization to further evolve the KM pyramid.

The knowledge sought and potentially included in the KMS discussed in this paper use the Jennex definition and postulate that a KMS must address traditional model elements and new ones identified and discussed by Jennex & Bartczak in 2013. The KMS of today must address vast amounts of data, a huge variety of content to generate value from data collected from many combined sources.

The Content Approach

The content of knowledge that must be stored in systems also has to be analyzed and may be described in terms of four core technological competencies that can deliver a competitive advantage to an organization. The content, as suggested by Leonard-Barton, includes skill and knowledge bases, physical technical systems, managerial systems, and values and norms of the organization (Barton, 1995).

In analyzing the content, some inroads have been made into the systems, but not in how they are to be constructed. For example, knowledge methods applied to the transport process that might move knowledge from one place to another within organizations have been categorized. Barton (1995) further suggests these methods include a recognized technical transfer capability (to a site) utilizing four approaches (assembly or turnkey, adaptation and localization, system, redesign, product design).

However, this development approach assumes the use of methodologies and tools that are not those employed in the information systems (much less KMS) of today, and further fails to suggest how the expanding data in the knowledge pyramid is to be captured and shared.

Enterprise Integration and Collaboration

Organizations that must later combine and share knowledge have followed the development trends of the day. The growth of database, data marts and data warehouses has driven this KMS effort to build enterprise KMS.

3. ISSUES IN IMPLEMENTING KM SYSTEMS

Quaddus & Xu (2005) point out the many concerns with KMS when they note that the long history of knowledge and knowledge management is over 4000 years old. They employed key personnel interviews and content analysis to identify factors and variables that impact KMS adoption and diffusion. The four major variables they identified affecting KMS adoption and diffusion were: organizational culture, top management support, benefits to individuals, and a dream of a KMS. However, they did not identify how the systems were developed or acquired or the issues involved in the broader acquisition process. According to Quaddus & Xu (1995) these acquisition issues, as well as the changing nature of information and the ways individuals and organizations adopt, use and defuse information with KMS introduce additional issues that impact the development process. Unfortunately, Quaddus & Xu's work stops short of discussing how these issues affect the development process.

4. KM DEVELOPMENT THEN AND NOW

Research shows that first generation knowledge management tools are based on knowledge being defined as explicit, with information portals leading to information supporting business processes. In second generation systems, Tzortzaki & Mihiotis, note that knowledge becomes dynamic as it constantly moves within the organization using four processes: socialization, externalization, combination and internalization. They further suggest that a third generation of KM systems is based on an emerging definition of knowledge as heuristic, or tacit, which requires the use and diffusion of this knowledge to be based on social processes (Tzortzaki & Mihiotis, 2014).

The traditional development approaches generally offered for information systems (but not explicitly called for in KMS development) is to match information systems with the organizational tasks to be supported or automated, thus ensuring usage and directly tying systems to organizational benefits and a ROI (Kankanhalli, Tan & Wei, 2005a, 2005b). The theory of task/technology fit (TTF) (Goodhue & Thompson, 1995) is a frequently followed approach that postulates that the use of technology is governed by establishing a match of equivalent fit between technology features and the demands of the user's task. It is founded upon the arguments that experienced users will rationally select tools and activities is they can accomplish their work with the greatest net benefit. A variety of technology utilization and adoption studies have been used to support this theory in various functional areas including accounting system adoption (Benford & Hunton, 2000), broad workplace use (Dishaw & Strong, 1999), online consumer shopping (Klopping & McKinney, 2004), and knowledge search repository usage for knowledge seeking (Kankanhalli, Tan & Wei, 2005a).

The methodology to build a KMS is described some 15 years ago by Tiwana, who sought to provide a methodology and instructional guide for constructing an enterprise KMS. This work describes hands-on techniques and tools for making a KMS, using existing intranets, data warehouses, and current project management approaches. A 10-step plan provides checklists to locate and audit the tacit knowledge you already have and maximize ROI from a KMS. It also identifies some of the limiting factors such as excessive formalization and overreliance on technology, and supporting master prototyping, and staffing with a Chief Knowledge Officer (Tiwana, 2004). However, the limitations previously noted do not go quite far enough in explaining why KMS are difficulty to construct. They do not explain the limiting and restrictive role of roles of technology in the development processes.

There are five reasons that logically suggest why LMSs are so difficult to develop. First, they are not true KMS, but are smaller, with targeted data and information sharing tools. Second, these tools are built to deliver decision related information (that may become knowledge) to distinct populations. They are therefore owned and supported by user communities, and not entire enterprises with strategic objectives in mind. Third, sharing methodologies and viewing mechanisms are designed for the targeted users – not broad populations. Fourth, platforms, servers, and database technologies are limited in extensibility and flexibility. Once systems are constructed, they are difficult to expand and enhance for these perspectives. Politically, they are owned, financially the costs are large to redesign, and finally, the technology underlying the systems may not easily be expanded. The organization may have multiple identity management systems that cannot be readily integrated to permit wide distribution of the data, information and knowledge. Finally, the systems are not designed for sharing – from two perspectives. The systems in today's world are subject to significant security risks and threats. Attacks come in many forms and may be internally and externally generated. This has prompted systems developed by offices or departments, and even those in large enterprises to protect the KMS investment with firewalls that prohibit most other organization members from accessing the KMS or even exploring the data resident there. It is a world of protectionism – with the many firewalls designed and maintained by disparate groups who are focused most clearly on protection, and not strategic sharing.

Today there are new techniques. The development processes today follow a Devops process that packages all the required software and delivers any number of advantages to the developer and supporting organization. Under this paradigm, containers incorporate all the files, configuration and environmental variables, and software needed to execute any application. The containers share resources, but do not need a complete operating system to run the applications. The container executes the image through an engine that deploys these images on hosts.

The technology is not necessarily new, and can use an open source engine and universal runtime. Competitors exist that perform similar functions, including some that use a separate engine relying on an open, standard container format. This technology utilizes micro services and distributed applications, and efficiently requires only limited resources from a host, since the containers operate independently.

Overall, the approach encourages flexibility and lets the developer design and implement non-standard images with new application libraries, because the developer only has to make changes to the code in the container image, and then can redeploy that image for a user. The high degree of flexibility can be understood because this technology is not the same as virtualization, where an operating system and application are permitted to only access the underlying hardware and resources through a hypervisor layer separating memory, compute and storage and the operating system and application and services.

Under this design and tool, applications run on their own version of an OS, and other applications on the same host may use different OS versions. Containers inside of virtual machines may have multiple OSes that are safe spaces for execution without interfering with other applications using the same OS (Tech Target, 2017).

With containerization there are gains in efficiency for memory, CPU and storage compared to traditional virtualization and physical application hosting, because there can be many more application containers on the same infrastructure. Application containers can run on any system, making them highly portable. Reproducibility is also high because the file systems, binaries and other information stay the same through the build, test and production cycle. Since version control is at the image level, configuration management is simplified. Scanners and monitoring tools are needed since containers are not isolated from the OS and security threats have easier access to the entire system. An organization must create policies to manage privilege levels for containers for security (Tech Target, 2017).

Given containerization, one can readily see the advantages for systems KMS that are facing the previously listed issues and deficiencies. First, flexibility, second, no need to rewrite, and third, improved logical access. Essentially, deploy your legacy systems, and develop a logical integration plan. Sharing can be cone with new applications and services (Tech Target, 2017).

The big change after the use of containers is that the KMS of the future can run in the cloud. The world today has moved to the cloud/ this move has greatly expanded and enable data sharing, and use of information systems. The cloud delivers well discussed benefits in terms of expandability, added storage, etc., but knowing the limits of KMS How does the cloud treat firewalls?

The essential concern with applying our understanding to the cloud and firewalls is to appreciate what firewalls do to protect an organization's network and users, and infrastructure and servers. Most organizations and users are familiar with firewalls stand-alone products or services designed to protect an enterprise network and its users. Firewall application may also act virtually to protect traffic going to, from, and between applications (in the cloud). They are not installed to protect a perimeter but to manage access inside public or private clouds and between/among applications.

Control is maintained through dashboards and management consoles that display activity and perhaps lets those select options permitting displays of information, blockings, etc. This can be extended to remote access users, connected via tunnels or VPNs. (Organizations may find savings in extra and unneeded firewall services, tools, as well as savings in OS licensing, and hardware using combinations of cloud and containerizations technologies. With this process, new zero-day threats or fixes, can be changed instantly. Thus avoiding need to download and install updates (Zeichick, 2017).

As an example, Microsoft's hybrid connectivity offers both Internet and network connectivity. This is effectively an extension of a tiered infrastructure via virtual networks (Ormond, Dial & Martin, 2017). Amazon's AWS approach is a similar security management service for rule configuration and management across your accounts and applications. Compliance is maintained with a common set of security enforced at the enterprise level in a consistent, hierarchical manner. It permits one to launch resources into a virtual self-defined network resembling a traditional datacenter network (Barr, 2018).

Understanding what cloud and containerization do for the enterprise KMS is essential. Understanding these technologies can lead to new uses of the old (smaller and targeted) KMS already in existence. Legacy applications need not be rewritten, lowering the costs. And importantly – the firewalls are not used for protection. In the cloud containerization environment, the networks are virtual and can be defined.

This suggests a strategy of planning, migration and integration that is supported by the evolving understanding of the knowledge pyramid. Not only does this synergy between knowledge and how it is used redefine KMS development methodologies, but continuing technology emergence further change the paradigms under which the applications are designed. For example, as previously mentioned, the emerging cloud environments not only provide an easy place to put legacy applications, they are actually changing the way that development is conducted. Previous constraints such as data storage, application size, and performance are no longer relevant due to continued streamlining of development and operational processes due to cloud technologies. As the knowledge pyramid and information system development capabilities continue to evolve, the need to continue migrating, integrating and understanding must continue apace.

5. KNOWLEDGE PROJECT IMPLEMENTATION

As far back as 1998, Davenport, De Long, & Beers provided a different approach to the development issue in their study of practical knowledge management development by studying thirty-one knowledge management projects in twenty-four companies. They saw these projects as attempting to use knowledge to support and meet organizational objectives. They recognized that the term *knowledge* may be difficult to apply in some of the projects studied, but that many have a limited impact. The project characteristics included all being unfinished, but having specific business and knowledge management objectives. The projects addressed knowledge, as opposed to information or data, and four broad types of knowledge objective (with one usually being primary): (1) creating repositories, (2) improving access, (3) enhancing the environment, and (4) managing the knowledge asset (Gandomi & Haider, 2015).

The analysis of these 4 types of knowledge projects is informative because it begins to identify the framework of how such KMS may be built. The repository projects studied by Davenport, De Long, & Beers (1998) store knowledge that can be collected or gleaned from items such as documents (containing memos, reports, presentations, articles that may have (or be) knowledge and holding them in a repository available for retrievable by to others, or where individual experiences can be reported and combined with others' comments. They found that (1) external environmental knowledge; (2) structured internal organizational knowledge; and (3) informal internal knowledge were stored as *lessons learned* or as raw information with an added context and synthesis that made it more understandable and accessible (valuable). Some systems also include specialized routing on different topics to those organization members with specific interest in a topic.

Davenport, De Long, & Beers (1998) also discuss the unstructured, and otherwise undocumented knowledge residing in the minds of the people in an organization that is commonly referred to as *tacit* knowledge elsewhere. They note that it is transferred from individuals and incorporated into repositories, through community-based electronic discussion that can spread tacit knowledge via sharing that previously occurred though organizational socialization processes.

6. TRANSFER PROCESSING

Ingesting knowledge was one thing. But Davenport, De Long, & Beers (1998) noted a second major type of project focused in dissemination by delivering knowledge or transferring it among individuals. This enabled others to obtain what the organization or other individuals knew and methods of sharing the knowledge through increased connectivity.

Today's KMS must also address the new forms of data, some of which were discussed earlier in the paper. This means they will have to include techniques (such as those of sentiment analysis for Big Data). However, the systems must be capable of adding advanced techniques because the tools previously used to analyze these data are not ideally suited to leverage Big Data where significantly more sentiment analysis will be required. (Gandomi, & Haider, 2015).

7. EXAMPLES

Research and documented description of how KMS are built is scant. Research data and efforts focus on questions of theory and the application of KMS to specific projects rather that enterprise wide solutions. How such systems are built remains undescribed. For example, Research and empirical evidence on how knowledge is managed in alliances has been discussed in an integrative and organized framework that illustrates how the knowledge management outcomes of knowledge creation, transfer and application are determined by four distinct sets of factors: knowledge characteristics, partner characteristics, partner interaction, and active knowledge management (Schiuma, Andreeva, & Kianto, 2012). But how this knowledge is created, retained, retrieved (in and from a KMS) and applied and how the interplay of the different factors affects knowledge management in strategic alliances remain widely unexplored according to these authors.

The 2010 Haitian earthquake complex response effort relied extensively on knowledge management systems (KMS) describes as social media technologies such as wikis and collaborative workspaces as the main knowledge sharing mechanisms. This example also points to the specific focus of a KMS with knowledge sharing, reuse, and decision-making features. It asserts that knowledge was maintained in these systems. However, important research questions remain unanswered regarding social media as knowledge management systems (Yates, & Paquette, 2011).

Examples of knowledge management systems available commercially are found via various search tools. For example, Captera provides a vendor directory, and a survey of KM products that broadly describes come of the features and benefits (deployment, categorization, collaboration, content management, full text search, knowledge base management, self-service portal). However, there is no data or background information to ascertain if the assembly of modules approaches the development issues identified, and the systems cannot be readily mapped to the literature research guidance and question posed in table 1.

8. ANALYTICS

This paper has discussed how the concept and context of knowledge management systems has evolved from explicit to tacit and social knowledge. This has changed the expectations on the outputs of KMS. The change from explicit to tacit knowledge has enabled use of Business Intelligence (BI) output to provide insight into business performance. It has also spawned use of Business Analytics, or using predictive models to provide support to business decision making (Pearlson, Saunders & Galetta, 2019). Use of data analytics continues to emerge, just as the context of knowledge as social capital continues to evolve.

It is intuitively obvious that future knowledge KMS leverage the new technologies and processes evident in analytics evolution to provide insight into the evolving social nature of knowledge. Machine Learning (ML) and Artificial Intelligence (AI) are becoming useful tools in providing answers to questions asked of KMS.

An example of ML used in this context is when the happiness of people on vacation is determined by the number of pictures of the inside of their hotel rooms posted on social media. The inference is that happy people are out having fun, while those enjoying their vacations less remain in their rooms.

As the social nature of knowledge matures, it is anticipated that AI applications will be useful in using this knowledge to provide insight. At present, there is little discrimination between AI and very good expert systems. However, as future AI systems improve in learning and synthesizing outcomes, it may be possible for AI applications not only to recognize patterns to predict outcomes, but also to independently determine variables, data sets and algorithms to provide answers.

Based on this discussion, it is imperative that future KMS incorporate increasingly complex analysis functionality to remain relevant. This is critical in analyzing and using social knowledge in the KMS environment.

8. CONCLUSIONS

The evolvement of KMS will continue as cognitive research advances our appreciation and understanding of how decisions are influenced by knowledge derived from these systems. Organizational learning, individual performance and uses of individual skills, accumulation and transfer processes will also change. A greater understanding of how different kinds of KMS provide knowledge that can be captured, stored, and accessed for organizational use and decision making is known. KMS refreshment and renewal priorities, and how KMS may replace and support complex and changing work activities must be developed. Despres & Chauvel suggest that KMS will continue to evolve, and predict that a new model of knowledge for the Theory of the Firm will elucidate new tactical values, principles, and judgments (2012).

As knowledge continues to evolve, emerging paradigms and technologies in analytics must be incorporated into KMS development to be able to use the increasingly social nature of knowledge to best advantage.

What is now understood of the ways knowledge is extracted and then employed from KMS is low. The theory of knowledge that is applicable to daily economic and applicable to business is not written or currently taught. How to build strong and renewable or defendable KMS within an organization is not well known. There is therefore much opportunity in examining and determining methodologies for developing and using KMS. There are many outstanding issues, or areas needing elucidation. Some of these are captured in the form of research recommendations in Table 1 below:

Issue for KMS Development	Literature finding/guidance	Research Recommendation
Wide Knowledge Content	Wide variety – static, dynamic. social	How to store and associate data types
Acceptance (deemed true)	Not addressed	Design formalized "certainty" process
Data – information – knowledge linage	Not Addressed	Hierarchical or indexing linking mechanism
Input - ingestion	Clean, filtering	Criteria maintenance
Processing	Sequences and steps; filtering; bi-directional	Acceptance; Bi-directional decision mechanisms
Success metrics	Meet "organizational goals"	Usage, access, value contribution
Project Approach	Structured	Compare to integration of current data and stores
Methodology	Hierarchical (step by step)	Comparison to Agile, DevOps
Environment	Cloud, hybrid	Transition, migration processes (for current knowledge)
Acquisition	Incremental, enterprise	Comparison – speed, acceptance, participation
Use, diffusion		
Dissemination process	Between individuals	Factors impacting use
Tools	Use existing now for target projects	Perform acer against new specialized KMS
Security	Internal firewalls, protection	Compare to cloud

Table 1. Research Recommendations

Associating KM with enterprise wide strategies, tactics, and operational activities may involve a deeper perspective of how individuals involves themselves with knowledge so they can constant learn and continually innovate. It is possible that KM systems may go beyond simplistic reasoning (without innovation) and deliver solid analyses of events and transactions by discovery of knowledge from database analytics. Intelligent agents must somehow then create and implement new actions that can become opportunities to provide services and positive results. The form that this will take cannot yet be predicted as knowledge as a social construct continues to emerge. One doctrine of KM is the need to arrange affairs to avoid rediscovering what earlier thinkers have created but maximize the reuse of valid knowledge and practices.

9. REFERENCES

Ackoff, R. L. (1989). From data to wisdom. *Journal of applied systems analysis*, *16*(1), 3-9.

Barr, J. (2018). "AWS Firewall Manager," Retrieved from https://aws.amazon.com/firewall-manager/

Barton, D. L. (1995). Wellsprings of knowledge: Building and sustaining the sources of innovation. Harvard Business School.

Benford, T. L., & Hunton, J. E. (2000). Incorporating information technology considerations into an expanded model of judgment and decision making in accounting. *International Journal of Accounting Information Systems*, *1*(1), 54-65.

Davenport, T. H., De Long, D. W., & Beers, M. C. (1998). Successful knowledge management projects. *Sloan management review*, *39*(2), 43-57.

Denning, S. (1998). What is knowledge management. *World Bank Knowledge Management Board, World Bank: Washington*.

Despres, C., & Chauvel, D. (2012). *Knowledge horizons*. Routledge.

Damodaran, L., & Olphert, W. (2000). Barriers and facilitators to the use of knowledge management systems. *Behaviour & Information Technology*, *19*(6), 405-413.

Dishaw, M. T., & Strong, D. M. (1999). Extending the technology acceptance model with task–technology fit constructs. Information & management, 36(1), 9-21.

Frické, M. (2009). The knowledge pyramid: a critique of the DIKW hierarchy. *Journal of information science,* 35(2), 131-142.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, *35*(2), 137-144.

Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. *MIS quarterly*, 213-236.

Hey, J. (2004). The data, information, knowledge, wisdom chain: the metaphorical link. *Intergovernmental Oceanographic Commission*, 26, 1-18.

Jennex, M. E. (2009, January). Re-visiting the knowledge pyramid. In *2009 42nd Hawaii International Conference on System Sciences* (pp. 1-7). IEEE.

Jennex, M. E. (2017). Big data, the internet of things, and the revised knowledge pyramid. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, *48*(4), 69-79.

Jennex, M. E., & Bartczak, S. E. (2013). A revised knowledge pyramid. *International Journal of Knowledge Management (IJKM)*, *9*(3), 19-30.

Kankanhalli, A., Tan, B. C., & Wei, K. K. (2005a). Understanding seeking from electronic knowledge repositories: An empirical study. *Journal of the American Society for Information Science and Technology*, *56*(11), 1156-1166.

Kankanhalli, A., Tan, B. C., & Wei, K. K. (2005b). Contributing knowledge to electronic knowledge repositories: An empirical investigation. *MIS quarterly*, *29*(1).

Klopping, I. M., & McKinney, E. (2004). Extending the technology acceptance model and the task-technology fit model to consumer e-commerce. *Information Technology, Learning & Performance Journal*, *22*(1).

Newman, L. (2008). Descartes' epistemology.

Ormond, J., Dial, J., and Martin, O. (2017). “Microsoft cloud services and network security,” Retrieved from https://docs.microsoft.com/en-us/azure/best-practices-network-security

Pearlson, K. E., Saunders, C. S., & Galletta, D. F. (2019). Managing and using information systems: A strategic approach. John Wiley & Sons.

Quaddus, M., & Xu, J. (2005). Adoption and diffusion of knowledge management systems: field studies of factors and variables. *Knowledge-based systems*, *18*(2-3), 107-115.

Ruggles, R. (1998). The state of the notion: knowledge management in practice. *California management review*, *40*(3), 80-89.

Sharma, N. (2008). The origin of the data information knowledge wisdom hierarchy. Data Inf. Knowl. Wisdom hierarchy.

Tech Target (2017). “Application Containerization,” Retrieved from https://searchitoperations.techtarget.com/definition/application-containerization-app-containerization

Tiwana, A. (2000). *The knowledge management toolkit: practical techniques for building a knowledge management system*. Prentice Hall PTR.

Tzortzaki, A. M., & Mihiotis, A. (2014). A review of knowledge management theory and future directions. *Knowledge and Process Management*, 21(1), 29-41.

Tuomi, I. (1999, January). Data is more than knowledge: Implications of the reversed knowledge hierarchy for knowledge management and organizational memory. In *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers* (pp. 12-pp). IEEE.

Wiig, K. M. (2000). Knowledge management: an emerging discipline rooted in a long history. *Knowledge horizons: the present and the promise of knowledge management*, *3*, 26.

Zeichick, A. (2017) "Understanding cloud-based firewalls," Retrieved from https://www.hpe.com/us/en/insights/articles/understanding-cloud-based-firewalls-1702.html

Understanding Campus Crime with A Multi-University Analytics System

Uche Iheadindu
MS Computer Science and Information Systems

Douglas Kline
klined@uncw.edu
Information Systems

Ron Vetter
vetterr@uncw.edu
Computer Science

Ulku Clark
clarku@uncw.edu
Information Systems

University of North Carolina Wilmington
Wilmington NC 28403

Abstract

Due to budget challenges, the campus police department of University of North Carolina Wilmington engaged in a data-driven performance management effort. To support this effort, publicly available data from multiple sources was integrated into rigorous data model in a single MS SQL Server database with interactive reporting using MS SQL Server Reporting Services. The data consisted of publicly available crime statistics for 38 universities, as well as characteristics such as Carnegie classification, acreage, budget, number of students, etc. The purpose of the system was to benchmark the campus police department against peer, aspirant, and other similar universities. This paper describes the architecture of the system, the benefits to the police department, and sample analytics.

Keywords: Performance management, data warehouse, campus crime, analytics, frontier analysis, visualization, database management.

1. INTRODUCTION

Like many organizations, the campus police department of University of North Carolina Wilmington faced challenges in budget and resource constraints. In order to make better management decisions about scarce resource allocation, leadership desired usable information upon which to make these decisions. The initial motivation was to benchmark performance of campus police against other "similar" universities. In other words, how is this police department doing, compared with others? However, "similar" can have many meanings:

- Similar in size?, e.g. # students, acreage

- Similar in setting?, e.g. urban, rural
- Similar in type?, e.g. private, public
- Similar in police challenges?
- Similar in budget?

All of these, it turned out, were factors that would be important to the benchmarking effort. Depending on the stakeholder, each of these might be important. No single source of data had all of these data items.

Secondary motivations for the system included being able to make informed data-driven decisions with limited resources, being able to make compelling arguments for additional resources, and perhaps being inspired by and learning from other institutions. Having good data opened possibilities that transcended the initial requirements.

This paper describes the creation of a database integrating data from multiple sources and pertaining to many universities, and the architecture of the system for extracting meaningful information from the data. This should guide developers and decision makers on the challenges encountered in applying analytics to a particular domain.

2. BACKGROUND

Campus Police Department

As at many universities, at UNC Wilmington safety is identified clearly as a goal in the strategic plan. In this way, campus police play a vital role to the organization. The campus police department must not only enforce city, state, and federal laws, but also support the mission of the University. University police are state certified law enforcement officers, carry arms, have full powers of arrest and have the same authority as other state police officers. They work closely with local police agencies. In addition to enforcement, the department educates the university community, and works with and guides the university administration.

The police department's most visible division is the Patrol Division. This is the largest in terms of staff and the most visible to the community. It operates 24 x 7 on foot, bicycles, and in patrol vehicles. In addition to deterrence and intervention, they also provide support such as security checks, car unlocking, and escorts.

The Investigation Division performs follow up on reported crimes, and cooperates with local, state, and federal agencies as necessary. They also provide expert insight to University leadership in matters of crime prevention, substance abuse, sexual assault, and awareness.

The Support Services Division primarily gather, record, and report data for compliance purposes. This division consists of sworn officers and tele-communicators. This division must record and report crime data according to the Clery Act (Government, 2019). This requires all institutions of higher education that have federal support to gather and report specified crime statistics. The resulting data is aggregated and published in the Integrated Postsecondary Education Data System (IPEDS) from the National Center for Education Statistics (NCES) (National Center for Education Statistics, 2012) The annual IPEDS reports included crime statistics as well as enrollment, budget, tuition, accreditation, and many other items.

The IPEDS report was the main source of data for this analytics system. Additional data was gathered from the US Census Bureau (US Federal Government, n.d.) and the Federal Bureau of Investigation (US Federal Bureau of Investigation, 2013). The Census Bureau data provided context for the Metropolitan Statistical Area (MSA), aka, community, in which the universities resided. The FBI data provided the crime statistics for the MSA. Finally, some budget data was gleaned from individual university web sites.

3. SYSTEM ARCHITECTURE

Requirements

Several data challenges presented themselves through the requirements. All data needed to be “citable”, with the ability to “drill-back” to the source of the data. Furthermore, data from these sources was aggregated across time, e.g., yearly data. Financial data was typically aggregated for the state fiscal year, enrollment data was aggregated for the academic year, and crime data was aggregated for a calendar year.

A second driving requirement was the need to be able to add new data sources, and new data items. For example, a local police department might also have data that could be integrated into the system in the future. The system had to be flexible enough to accommodate integrations with other data sets in the future.

Finally, the system had to be dynamic, with the ability to interact with the data. In the past, information was presented in periodic static reports. An annual report is quite stale even at the moment of publication. Also, the report is limited to what the authors decided to include. The police department wanted domain-expert police personnel to be able to explore the data interactively.

Platform Selection
The existing data processing occurred in long-lived and ad hoc spreadsheets, and in a home grown single-user Access database. This effectively created data silos and limited the utility of the data. Data quality was a concern due to manual, questionably repeatable data manipulation, and a lack of data integrity checks. Versioning of reports, spreadsheets, and the Access database had become problematic. Backups were ad hoc, manual, and had no consistent policy driven automation.

The decision was made to move to a more enterprise-level architecture. The desired architecture would be a fully transactional relational database server with a reporting server. MS SQL Server was chosen with SQL Server Reporting Services, both to be hosted and managed by the university technology services division. This ensured that servers were being monitored, that data was being backed up along with organizational policy, that the servers were in a data center with redundancy, and that the systems were robust to personnel changes.

Data Model
This system demanded a custom relational model to address the requirements. Central to the model was the ability to have many different types of metrics. Also required was the ability to add metrics as they became available. Figure 1 shows the data model for handling the different types of metrics and their values.

In this model each MetricValue represents a single data item, e.g., the number of sworn officers for a particular university obtained from a particular source, for a certain year with certain year type (yID). The current system has 208 Metric records, with 16,678 distinct MetricValues. Note that each Metric subtyped as a CrimeMetric, ControlMetric, or DiscMetric (discretionary). Discretionary Metrics included items such as budget, number of patrol officers, etc. Control Metrics are items like population of the MSA surrounding the university. Note that the use of subtypes is necessitated mainly by CrimeMetrics being related to other CrimeMetrics.

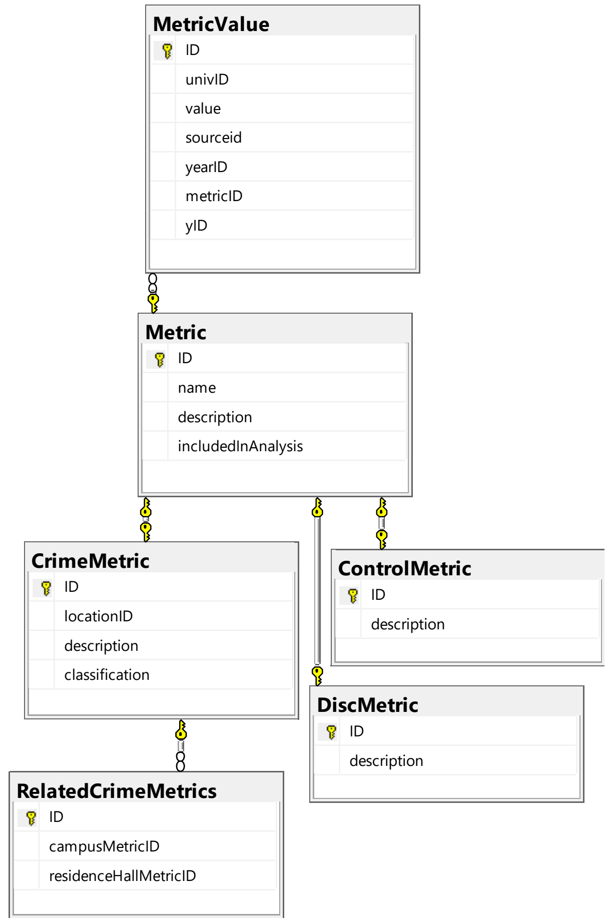


Figure 2: Metric Data Model

The data model in Figure 1 shows how data related to universities is stored. Figure 2 shows how Metropolitan Statistical Area (MSA) data is stored and related to universities. A single MSA can be the home of multiple universities. A refinement of this data model might have MetricValue as the supertype of a UniversityValue and MSAValue, since MetricValue and MSAValues are very similar except for foreign keys to University and MSA, respectively.

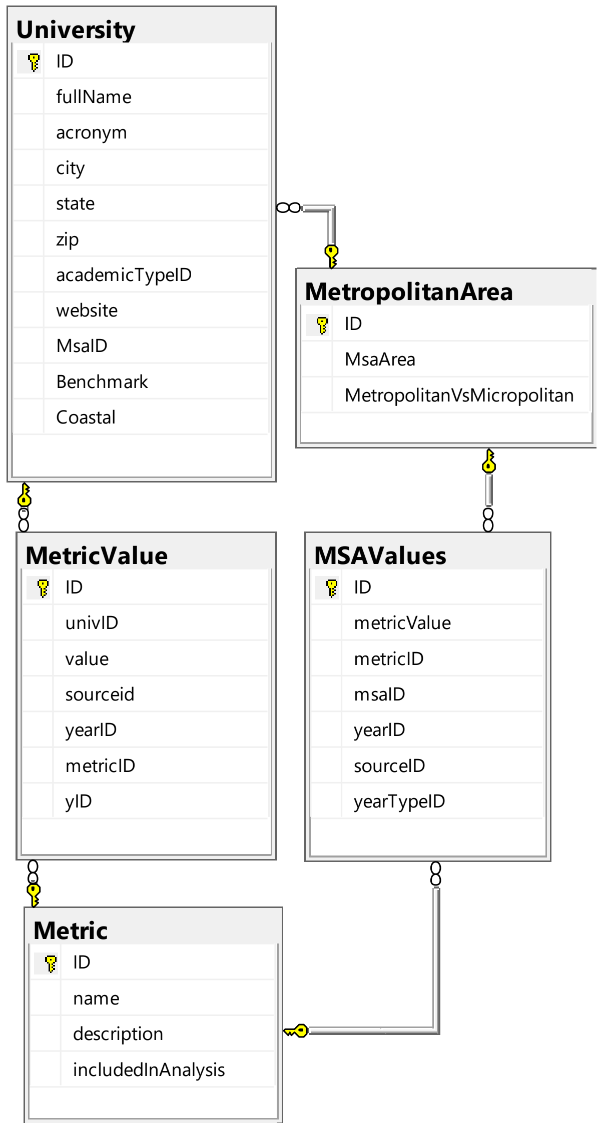


Figure 3: Metric Values Data Model

Crime metrics were quite complicated, with many "qualifications" on each metric. Figure 3 shows the data model for crime metrics. Note that CrimeMetric, also shown in Figure 1 is a subtype of metric, and that HateCrime is a subtype of CrimeMetric, i.e., some crimes are hate crimes and have additional data. Crimes can also be related to other crimes, so CrimeMetric has a many-to-many relationship with itself via RelatedCrimeMetrics. In Location, the location field is actually a location type with these values: Campus, Residence Hall, and OffCampus.

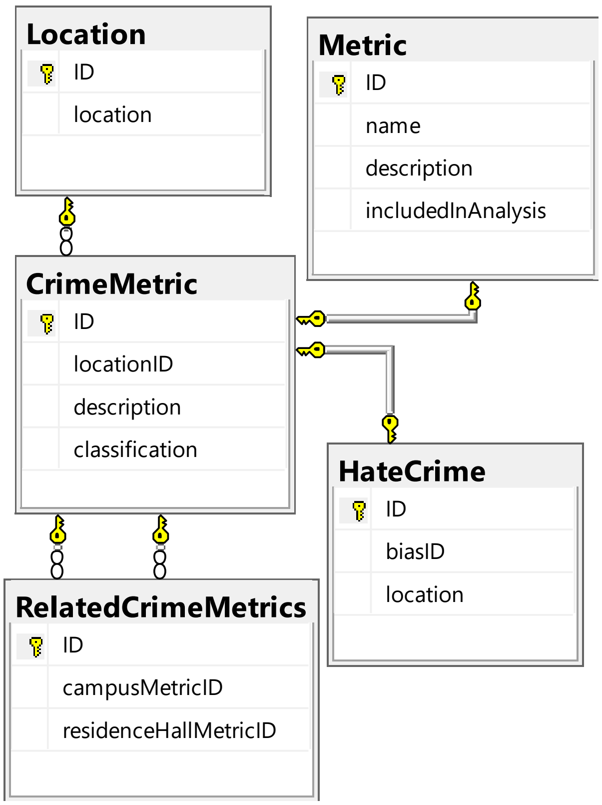


Figure 4: Crime Type and Location Data Model

Reporting Architecture

To better manage the complexity of necessary reporting capabilities, and to present a simpler mental model to users, we extensively used views. Figure 4 shows the flow of data from database tables to finished reports.

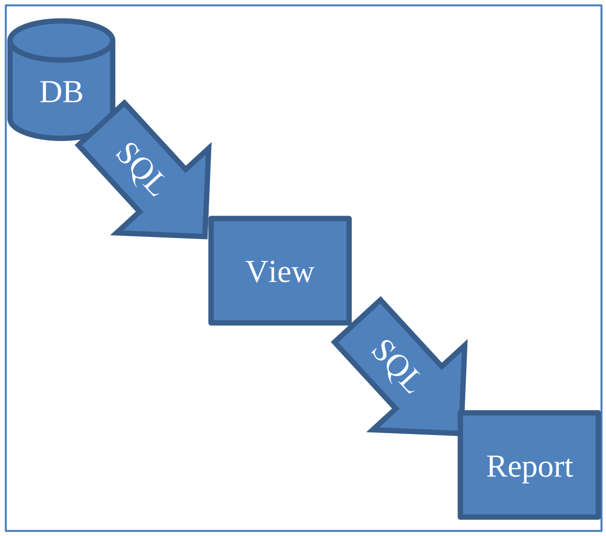


Figure 5

The first SQL statement in the process mainly joins tables appropriately. This eliminates the need for a report writing user to understand the relatively complex data model. For example, a view dealing with CrimeMetrics might need a self-referencing many-to-many outer join. Having a single View that performs this not only promotes code re-use and modularity but saves the report-writer time and potential mistakes.

The second set of SQL in the process might reside in the database server, or in the report itself. This SQL likely limits records with a WHERE clause or aggregates with GROUP BY. In this system, we used parameterized queries, where the parameter values were set via a drop-down or text field in the report and sent as parameters to the SELECT statement.

This design decision to use views was important but can be controversial. Views can have an impact on performance and with heavily nested (a view references another view) use can make the SQL code hard to read. In our situation, the size of the data is not likely to create a performance issue. The readability of the many views must be managed with naming conventions and very descriptive names. The current system has 123 views.

Reporting, Analytics, Visualization

In total 73 reports were designed for various purposes, and deployed using SQL Server Reporting Services (SSRS), where authenticated users could access them via a web page and click to download the data in .xls format. Simple tabular and graphical reports can be created by police department staff through the web-based report designer appropriate for technical novices. These can be saved and shared on the server for future use. An example of a repeatable tabular report might be the number of offenses per year for each university.

We also added interactive functionality where, for example, offenses per year could be drilled-down into counts of types of offense or clicking on a university drilled-through to a university-specific report. Many of the reports featured standard visualizations such as Pie-charts, bar charts, etc.

A particularly helpful interactive visualization is the scatterplot of universities on choose-able metrics. Here, the x-axis and y-axis are selectable from all metrics, and all universities are presented as dots in a scatterplot labeled with the university names. This allowed comparison of universities on any pairwise combination of metrics. This was very interesting to those knowledgeable in the campus crime domain, though perhaps unintelligible to the casual observer. For example, setting campus acreage on the y-axis and number of residence hall beds on the x-axis immediately showed how universities related to each other in a student-density sense. Then changing the x-axis to alcohol violations and back again gave a crude interactivity. This was highly engaging and had much more meaning to knowledgeable police professionals than to technical and analytics professionals.

The system was helpful for the benchmarking purpose. Although universities identify academic peer, sister, and aspirant schools, there is not an equivalent with respect to campus police challenges. Through interacting with the data, the police department was able to identify this subset of metrics to use in identifying universities with comparable crime challenges:

- Student headcount
- Total dormitory capacity
- Employee headcount
- MSA Population
- Total Acreage
- Total Operating Budget Expenses
- Operation and Maintenance Plant Expenses

Student headcount and employee headcount are measures of the population at the university. Total dormitory capacity indicates the proportion of residential population. The MSA population is a rough indicator of the urban/suburban/rural setting of the campus. Total acreage effective defines the area requiring monitoring. Total Operating Budget Expenses gives a rough measure of the total University budget, while Operation and Maintenance Plant Expenses (where Campus Police are generally placed) is an indicator of the police department budget. Using a normalized Euclidean distance measure, a set of six universities were determined to be very similar to UNC Wilmington.

Other analytic techniques were explored with limited success. An attempt was made to correlate discretionary metrics with crime metrics, e.g., number of sworn officers (discretionary) and number of alcohol violations. Unfortunately, the discretionary metrics are not available through the IPEDS data sets, and thus required surveying a subset of universities, resulting in only 6 universities participating. In the end, correlations were not significant at the alpha = 0.05 level.

Data Envelopment Analysis (Anderson, Sweeny, & Willimas, 1994) was then performed on this subset of similar universities. This technique gives a way to examine “efficiency” through comparison of discretionary metrics with outcomes (crime metrics). This analysis was foiled mainly by too few universities (6) with respect to number of metrics. This caused almost all universities to appear 100% efficient. Through this analysis, we also realized a domain-specific challenge: relationships between discretionary metrics and outcomes are complex and commonly non-linear. For example, adding more patrol officers increases tickets written linearly to a point, then adding more officers makes tickets decline. With few officers, the crime rate is constant, and additional officers merely *detect* more crime instances, resulting in more tickets. The presence of more officers begins to have a *deterrent* effect, actually decreasing the crime, and resulting in fewer tickets.

6. CONCLUSIONS

Design, construction, and use of this system was extraordinarily instructive, and useful to the campus police department. The main goal was met: to determine a data-driven approach to identifying benchmark universities with respect to campus police department challenges. The platform can be extended with user-authored reports as needed.

One glaring limitation is that the data loading was not scripted in such a way to be perfectly repeatable. Assuming the IPEDS, FBI, and US Census data report formats remain consistent, a scripted Extract, Transform, Load should be easily achievable with modern data manipulation tools.

In the future, we hope to add more data, mainly through annual IPEDS reports. With data sharing among the similar universities, it would also be possible to gain enough data to appropriately use statistical tests and techniques such as Data Envelopment Analysis.

7. REFERENCES

Anderson, R. D., Sweeny, D. J., & Willimas, T. A. (1994). *An Introduction to Management Science - Quantitative Approaches to Decision Making.* New York: West Publishing Company.

Government, U. (2019). Federal Code of Regulations. *EDUCATION*. Retrieved 7 15, 2019, from https://www.law.cornell.edu/uscode/text/20/1092#f

Iheadindu, U., Kline, D, Vetter, R., Clark, U., Donaldson, D. (2013) A Multi-University Data Analytics System for Understanding Campus Crime. Annals of the Master of Science in Computer Science and Information Systems at UNC Wilmington, 7(2) paper 6. http://csbapp.uncw.edu/data/mscsis/full.aspx

Johnston, T., & Weis, R. (2010). Managing Time in Relational Databases. San Francisco: Morgan Kauffman Publishers.

Larson, B. (2012). Microsoft SQL Server 2012: Reporting Services. New York: McGraw-Hill.

Microsoft. (2012, 10 3). Understanding Simple and Complex Expressions (Reporting Services). Retrieved from Microsoft: https://docs.microsoft.com/en-us/previous-versions/sql/sql-server-2008-r2/ms345242(v=sql.105)

Microsoft. (2013, 7 31). Authentication with Report Server. Retrieved from Microsoft: https://docs.microsoft.com/en-us/sql/reporting-services/security/authentication-with-the-report-server?view=sql-server-2017

Microsoft. (2013, 7 31). Working with Data in Reporting Services. Retrieved from Microsoft.com: https://docs.microsoft.com/en-us/previous-versions/sql/sql-server-2005/ms160324(v=sql.90)

National Center for Education Statistics. (2012). IPEDS Data Feedback Report. National Center for Education Statistics. Retrieved from https://nces.ed.gov/ipeds/DataCenter/DfrFiles/IPEDSDFR2012_199218.pdf

Russ, M., Ferrari, A., & Webb, C. (2012). SQL Server 2012 Analysis Services - The BISM Tabular Model. Sebastopol: Microsoft Corporation.

US Department of Education Office of Postsecondary Education. (2013). The

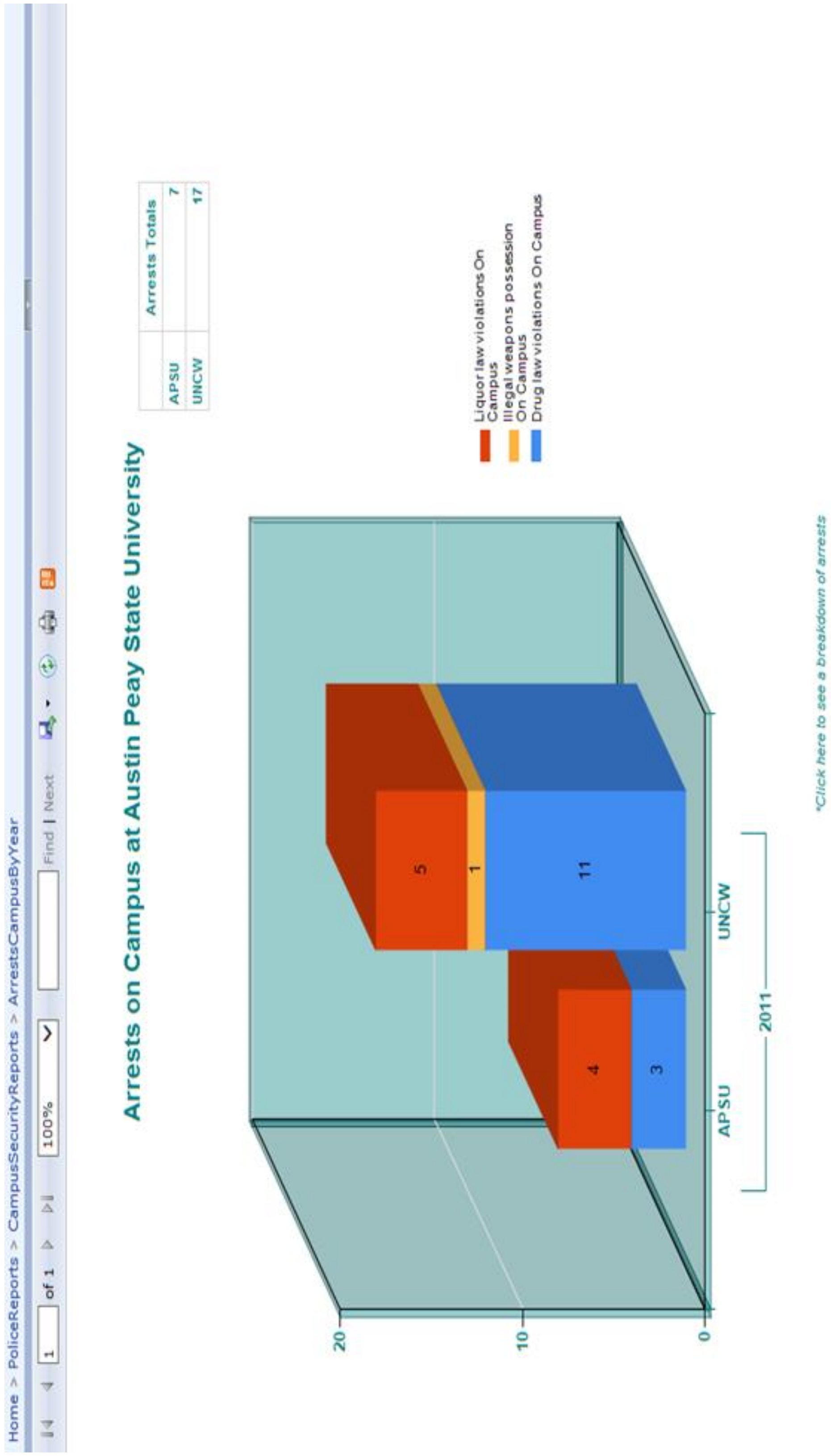
Handbook for Campus Safety and Security Reporting. US Department of Education. Retrieved from https://www2.ed.gov/admins/lead/safety/handbook.pdf

US Federal Bureau of Investigation. (2013). Uniform Crime Reporting Statistics. Retrieved from https://www.ucrdatatool.gov/

US Federal Government. (n.d.). US Census Bureah. Retrieved from https://www.census.gov/

Yeruva, S., Kumar, P., & Padmanabham, P. (2012). Distributed Warehouses: A Review on Design Methods and Recent Trends. International Journal of Computers and Distributed Systems, 60-61.

Appendix 1


Home > PoliceReports > CampusSecurityReports > ArrestsCampusByYear
1 of 1
100%
Find | Next
Arrests on Campus at Austin Peay State University
Arrests Totals
APSU 7
UNCW 17
Liquor law violations On Campus
Illegal weapons possession On Campus
Drug law violations On Campus
20
10
0
4
3
5
1
11
APSU
UNCW
2011
*Click here to see a breakdown of arrests

Trash to Treasure: Predicting Landfill Gas Flow to Optimize Electricity Generation

Edgar Hassler
hassleree@appstate.edu

Dan Emery
emeryde@appstate.edu

Jason Hoyle
hoylejw@appstate.edu

Joseph Cazier
cazierja@appstate.edu

Center for Analytics Research and Education
Appalachian State University
Boone, NC 28608-2049, U.S.A

Abstract

Data analytics and machine learning have the potential to modify and improve many old school businesses. Among the oldest businesses for the human race is managing the waste we generate. In this paper we show how data science can be applied to help derive increased value from a byproduct of that waste, landfill gas. Gas produced from the decomposition of waste in landfills can be captured and transformed into a resource that benefits the local community, environment, and economy. We use analytics to better understand how weather conditions impact the methane content of landfill gas in ways significant enough to interfere with its use as a source of energy. We model methane concentrations in landfill gas and use machine learning techniques to predict future changes in methane concentration using a database of weather, water composition, and landfill gas collection performance metrics. A multilayer predictive model of methane concentration is developed that will aid in the transformation of day-to-day operations of landfill gas collection to maximize the utilization of gas extracted from the landfill, while minimizing the cost of pollution mitigation. This can help transform the industry while mitigating some environmental concerns.

Keywords: Sustainability, Landfill Gas, Analytics, Machine Learning, Methane Prediction

1. INTRODUCTION

Current global municipal solid waste generation levels are approximately 1.3 billion tonnes per year and are expected to increase to approximately 2.2 billion tonnes per year by 2025 (Hoornweg & Bhada-Tata, 2012). Waste that is not recycled or incinerated is placed into landfills where its decomposition releases methane gas and carbon dioxide into the atmosphere accounting for an estimated 17.6% of human-related methane gas emissions in the United States (EPA, 2017).

In addition to climate change impacts, methane and other gases created in a landfill pose risks to air quality as well as human health and safety. To limit these impacts, emissions from landfills are controlled by installing gas collection systems and either burning the landfill gas (LFG) in a flare or utilizing it as an energy source to produce electricity, steam, or even vehicle fuel. Using LFG as an energy source transforms post-closure management of landfills from a cost center into a profit center and has the potential to transform foundational thinking in the waste management industry from viewing trash as waste into viewing trash as a resource—after all, it's only waste if it's not used.

In order to make the gaseous byproducts of our waste more useful, there are several analytics techniques that have the potential to help. These include

- **Modeling –** In order to apply analytic techniques to help convert this waste to useful energy, we must first model and understand the drivers of gaseous levels and concentration. This requires discovery and weighting of key drivers.
- **Prediction –** Next, as the relationships are better understood though discovery, we can apply that understanding by building predictive models based on key variables to forecast what is likely to happen next. This is critical to building a useful solution.
- **Automation –** Finally, as the relationships are understood and can be reliably predicted, these can be incorporated into an automatic solution using information and robotics technologies.

These stages of data analytics to model a current relationship, predict a future one and automate a solution are critical to the use of data science for building usable solutions with real impact.

For example, regulations requiring the monitor and control of LFG emissions specify a minimum frequency of monthly visits (Monitoring of Operations, 2000), and these requirements mean a technician must physically visit each of hundreds of gas wells to take measurements using handheld instruments and make manual adjustments for decades after the landfill closes. This infrequent, labor-intensive, and time-consuming manual process is not only an imprecise, potentially arbitrary, approach to managing LFG collection for the control of emissions, but it also produces an LFG supply that is highly variable in both gas flow and methane concentration that limits the value of using this energy-rich gas as fuel. **Automation** can add real value here, but first it must be modeled and predicted reliably.

In terms of **modeling** there is a considerable body of research on the production and diffusion of methane gas in a landfill (Rachor et al. 2013, Alexander 1971, Spokas et al. 2006, Farquhar & Rovers 1973, Bade Shrestha & Narayanan 2008, Toerien & Hattingh 1969, among others); an expanding literature on approaches and strategies to maximize LFG generation, extraction, and use as fuel (Reinhart 2002, Reinhart 1996, Townsend 2018, Ozkaya 2007, Warith 2003, Buivid 1981, Kinman 1987); and a long line of research into the internal dynamics of landfills that influence methane generation emissions, migration, and extraction such as pressure, cover permeability, depth, moisture content, waste composition, temperature, etc. (Arigala et al. 1995, Chen et al. 2003, Xi & Xiong 2013, Hashemi et al. 2002, Sanchez et al. 2006).
On example of a useful model using the fuzzy logic algorithm was derived by Garg et al. (2006) where they built a useful model based on average climate and waste site characteristics.

While these studies and others of a related nature contribute to increased knowledge of methane-related processes inside a landfill, the few that even address time do so on scales of months, years, or decades.

We need **prediction** however, when using LFG as fuel, particularly in a continuous process such as generation of electricity or steam, where changes in the performance of the collection system over periods of months or years are significantly less important than changes in collection system performance over *time intervals of minutes or hours*. Significant changes in methane concentration and, to a lesser extent, changes in gas flow rate will adversely impact the operation, efficiency, and output of LFG-fueled energy conversion devices (e.g. boilers or electricity generators) and at times even cause shutdown events. There exists a limited body of research examining the causes and nature of variability in LFG emissions or gas collection system performance over short time intervals (e.g. hourly or sub-hourly) and few, if any, efforts that attempt to forecast the near-term performance of a landfill gas collection system on an hourly or sub-hourly basis as we have done here.

Absent external influences, landfill gas pressure will stabilize at a landfill-specific level, and once steady-state conditions are attained then gas emission or migration out of the landfill occurs at the rate of gas generation. Lu and Kunz (1981) demonstrated that extracting LFG lowers the gas pressure during the time the vacuum is applied, but when the vacuum is removed the internal gas pressure quickly returns to its steady-state level, implying that gas extraction functions as a substitute for gas emissions and that factors impacting gas emissions will also impact gas extraction.

Young (1990) described the inverse relationship between changes atmospheric pressure and the volume of LFG emissions concluding that the amount by which emissions change over a sub-hourly time interval is proportional to rate of change in atmospheric pressure rather than the actual value of atmospheric pressure. In a further study, Young (1992) explained several factors that cause carbon dioxide and methane – the two largest constituents of LFG – to respond differently to changes in atmospheric pressure. In particular, he demonstrates, in stepped time intervals of less than one-half hour, methane concentrations rising from 50% to 55% over a period of hours and then rapidly and consistently falling to about 40% over the next day or two in response to changing atmospheric pressure.

Aghdam, Scheutz, and Kjeldsen (2019) summarize the conflicting state of conclusions reached in some of the limited number of empirical studies examining the influence of meteorological factors on LFG emissions or collection system performance. Among those factors, absolute level of barometric pressure, rate of change in barometric pressure, solar radiation, wind speed, soil moisture, air temperature, and soil temperature have been found to both have a significant influence and also not to have a significant influence on gas volume and composition.

The inconsistency of results among empirical studies coupled with the complexity resulting from the multitude of interrelated factors that may influence LFG flow and composition over short time intervals provides an opportunity to apply modern data science techniques to improve the performance of LFG energy systems. By carefully analyzing the application and data characteristics, researchers and practitioners can then adopt or develop the appropriate analytical techniques to derive the intended impact (Chen et al. 2012).

Machine learning and applied analytics have the potential to transform the way in which this one-time pollutant is transformed into both a valuable energy source benefiting local economies and a new source of value for the waste management industry, while also mitigating its environmental impact. Analytics also has the potential to automate the current costly process of collection system tuning, identify equipment failures and needed repairs much more rapidly, and greatly improve safety at and around the landfill site by remotely identifying the presence of hazardous gases such as methane and hydrogen sulfide at the landfill surface.

As the economic and environmental benefits of LFG use are increasingly recognized, data analytics and machine learning can aid in the optimization of LFG collection and spur a transformation in the waste management industry. By maximizing the amount of LFG collected and selling the recaptured gas, energy, or carbon credit, a landfill can literally turn trash into treasure.

Research Objective

As gas is extracted from the landfill, methane concentration of the gas fluctuates drastically. If methane concentration drops too low, the gas becomes unusable as an energy source, induces stress on the generating engine and can lead to damage or failure of the engine.

By using data analysis techniques, this project specifically aims to model methane concentration in near-real-time and thus enable landfill operators to predict future changes in methane concentration and gas flow to preemptively respond to methane concentration changes by taking the necessary actions to protect the engine, keep LFG composition optimal, and maximize profit while minimizing the negative effects on the environment.

The remainder of this paper is organized as follows. Section 2 provides background on the business case, environmental impacts and foundational knowledge of LFG production. Section 3 describes the methodology of the study including the dataset development and model construction. In Section 4, the results of the modeling process are presented followed by discussion in Section 5. Section 6 concludes the paper and notes possible future directions for research.

2. BACKGROUND

Business Case
The high concentration of methane gas in LFG incentivizes landfill operators to capture the methane gas diffusing out of the landfill for reuse or sale. A study based on 8,196,000 tonnes municipal solid waste generated in Peninsular Malaysia in 2010 estimated that the collection and sale of landfill gas generated 1.9 billion kWh of electricity per year, worth over US$190 million (Johari et al., 2012). Another study found that capturing the methane emissions in Oman from 2016–2030 will attract a total revenue of approximately US$333 million from carbon reduction, and approximately US$291 million from electricity generation (Abushammala et al. 2016). With such money at stake, the importance of understanding more about LFG production and composition in a landfill is heightened.

Currently, the Environmental Protection Agency's Landfill Methane Outreach Program has identified 634 operational landfill gas collection projects in the United States. Combined, these projects generate 291 mmscfd (million standard cubic feet per day)—enough energy to supply the state of Maine's annual natural gas consumption. If sold at $3.50/1000 scf, it would generate US$1,018,500 in revenue every day solely from the sale of the collected gas.

Landfill gas is also a cheap source of energy. Local businesses realize cost savings associated with using LFG as a replacement for more expensive fossil fuels. Some companies could save millions of dollars over the life of their LFG energy projects (LMOP 2017). General Motors has leveraged this cheap energy source into massive savings for the company. A statement from General Motors Green on 13 March 2016 reads, "A decades-long approach to sourcing renewable energy has produced lessons learned that helped GM further reduce its environmental footprint and save $80 million along the way." This testimonial serves as a pilot-study of sorts as to how much the utilization of landfill gas as an energy source can benefit a company.

While the energy savings and environmental benefits of using LFG are potentially substantial, maximizing the value of these benefits requires effective management of the LFG collection system as even short-duration system shutdowns can result in significant financial losses. For example, a landfill selling 2 mmscfd at a price of $3.50/1000 scf loses $7,000 in revenue for every day the system is not operating.

Environmental Impacts
As stated previously, landfill gas is a byproduct of society. It is important to note this paper is not suggesting that collecting and combusting LFG eliminates the need to decrease waste production. It is simply pointing out that capturing and combusting LFG will always be more environmentally friendly than allowing the gas to escape into the atmosphere.

Methane gas (CH_4), has a molar mass of 16.04 g/mol. Carbon Dioxide (CO_2), has a molar mass of 44.01 g/mol. The combustion of methane gas converts 1 mole of methane gas to 1 mole of carbon dioxide gas. Thus, combusting 1 ton of methane gas leads to the release of 2.74 tonnes of carbon dioxide gas.

Previous studies have shown that methane gas has a 20-year global warming potential of 86 when considering carbon-climate feedback (Myhre et al. 2013). Carbon dioxide is used as a baseline for global warming potential values, giving it a value of 1.

Using these global warming potentials, we assert that each ton of methane gas combusted reduces the 20-year global warming potential of the gas by 83.26. Essentially, 1 ton of methane gas released to the atmosphere carries a global warming potential of 86. By converting 1 ton of methane gas to 2.74 tonnes of carbon dioxide, the resultant gas carries a global warming potential of 2.74, a ~97% reduction in global warming potential.

Landfill Gas Collection Systems
Landfill gas collection systems are remarkably complex systems, featuring anywhere from dozens to hundreds of interconnected gas wells located throughout the landfill, all applying a vacuum to extract the gas produced by myriad chemical and biological processes underway within the landfill.

These components all interact with external factors, such as weather, and the resulting gas extracted at a given moment varies significantly in volume and composition. Yet, despite this complexity, LFG collection systems are typically managed by a technician who takes measurements using handheld instruments and adjusts the collection system by turning valves on a largely ad hoc basis.

Using applied analytics techniques to understand and predict the underlying biochemical processes and physical factors that impact LFG quality will enable a far more efficient, cost-effective, and profitable use of LFG as an energy source, while simultaneously increasing environmental and operational benefits.

Processes Within the Landfill

When municipal solid waste is first deposited in a landfill, it undergoes an aerobic (with oxygen) decomposition stage where little methane is generated. Then, typically within less than one year, anaerobic (without oxygen) conditions are established and methane-producing bacteria begin to decompose the waste and generate methane. Thus, landfill gas production can be broken down into two major stages, the non-methanogenic stage and the methanogenic stage (See Figure 1).

The optimum temperature range for aerobic decomposition is 54 to 71ºC, while the optimum temperature range for anaerobic bacteria is 30 to 41ºC. A dramatic drop in activity of anaerobic bacteria has been noted at temperatures below 10ºC (USACE 2013).

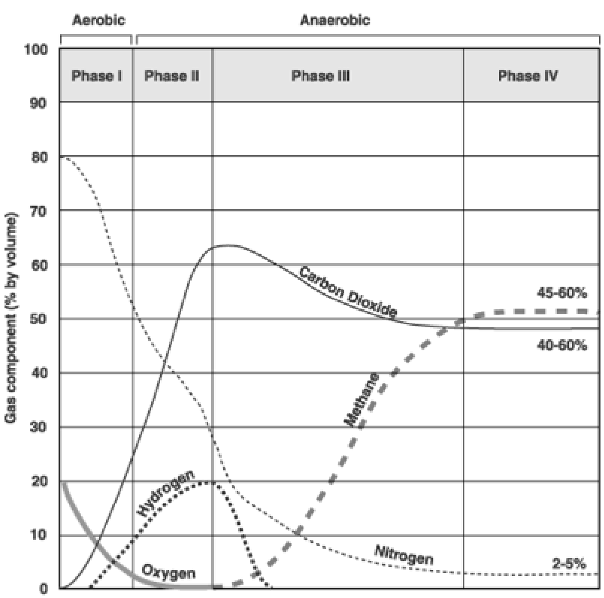


Figure 6. Landfill gas composition over time (EPA, 1997). Phase duration time varies with landfill conditions.

Precipitation dramatically affects the LFG generation process by supplying water to the process and by carrying dissolved oxygen into the waste with the water. High rates of precipitation may also flood sections of the landfill, which will obstruct LFG flow. The amount of precipitation that reaches the waste is highly dependent on the type of landfill cover system (USACE 2013).

Molecular diffusion occurs in a system when a concentration difference exists between two different locations. The concentration of a volatile constituent in the LFG will almost always be higher than that of the surrounding atmosphere, so the constituent will tend to migrate to the atmosphere. Wind often serves to keep the surface concentration at or near zero, which renews the concentration gradient between the surface and the interior of the landfill, thus promoting the migration of vapors to the surface. Geomembranes in landfill covers significantly reduce diffusion as the geomembrane prevents LFG from diffusing to the atmosphere (USACE 2013).

Advective flow occurs where a pressure gradient exists. The rate of LFG movement is generally orders of magnitude faster for advection than for diffusion. LFG will flow from higher pressure to lower pressure regions. In a landfill, advective forces result from the production of vapors from biodegradation processes, chemical reactions, compaction, or an active LFG extraction system. Variations in water table elevations can create small pressure gradients that either push gases out (rising tide) or draw gases in (falling tide). Changes in barometric pressure at the surface can also have an impact on the advective flow of LFG (USACE 2013).

3. METHODS

Data for this project was collected from a midsize rural landfill in the Southeastern United States.

Landfill Monitoring System

Landfill gas flow and methane content are continuously recorded by an Eurotherm 6180A Paperless Chart Recorder. A Thermal Instrument Model 9500 flow meter was used to correct for temperature and pressure at 60 degrees Fahrenheit at 1 atmosphere of pressure. Gas was analyzed using the Landtec FAU-TDL gas analyzer.

Field check calibration for all equipment was performed. The project was verified to the Climate Action Reserve's Landfill Project Protocol Version 3.0. All data was collected prior to the beginning of this research.

Having these systems in place should allow for an automated system to be implemented with little changes to the current setup, which should ease the deployment of the model.

Dataset Description

Observations were recorded every 20 minutes from midnight 14 June 2012 to 5:00 pm 31 August 2013. The initial dataset contained 27,962 observations and 38 variables. Of the 38 variables, 12 describe weather conditions. The weather data was collected from a nearby rural airport weather station.

The 12 weather variables measure six different weather metrics. Each observation recorded the six metrics at the local airport and provided a regional analysis. The dataset also included 25 landfill gas collection system performance metrics. A timestamp of each observation was also included.

There are periods of missing data, half of which occur within a 30-day period in 2012 and during a couple weeks in Feb 2013. The missing time periods for the landfill gas collection system performance metrics are as follows:

- 7/10/2012 - 8/18/2012
- 10/09/2012: 6 hours
- 10/10/2012: 6 hours
- 10/22/2012: 7 hours
- 02/13/2013 - 02/19/2013
- 07/09/2013 - 07/17/2013

There were periods of missing values for the weather data as well. Exploratory analysis of these missing values showed that if one variable was missing, then all values were missing. The time periods never lasted longer than two hours. The short timeframe of missing values for the weather data allowed for the values to be imputed using the following equation:

$$y_i = \frac{(y_{i-1}^* + y_{i+1}^*)}{2} \qquad (2)$$

where y_i is the missing value, y_{i-1}^* is the previous recorded instance of that variable, and y_{i+1}^* is the next recorded instance of that variable.

Feature Construction

Precipitation was not included in the initial weather dataset. Knowing that moisture plays a large role in landfill gas production, precipitation information was acquired through Climate Data Online. Daily precipitation values were joined to the existing dataset by the date of the observation.

Water quality information was obtained from a dataset made available by the United States Geological Survey for a river approximately 50 miles southeast of the county landfill. While not an optimal indicator of conditions on the site, it is included to provide, at a minimum, a regional indication of water quality due to lack of instrumented reading at the facility. Observations were recorded every 15 minutes. The primary variables of interest in this dataset were water elevation, dissolved oxygen in the water, and pH of the water. Prior to joining the datasets, the water quality data was grouped by date and hour of day. Following this step, average elevation, dissolved oxygen, and pH was calculated. The two datasets were then joined by date and hour of the day.

One common thread between all landfill gas models is the importance of landfill moisture. Access to weather data will allow for more information on moisture. The three and five-day cumulative precipitation values were calculated from the daily precipitation values.

$$3\ day\ precipitation\ for\ row\ i\ = \sum_{j=1}^{3} P_{i-j} \qquad (3)$$

$$5\ day\ precipitation\ for\ row\ i\ = \sum_{j=1}^{4} P_{i-j} \qquad (4)$$

Oxygen content in the landfill is yet another obvious driver of methane gas production. As mentioned previously, the primary methane-producing reaction in a landfill is anaerobic. The dissolved oxygen content of a nearby river serves as a proxy for dissolved oxygen in the water present in the landfill. The problem with this variable is that dissolved oxygen can be high, but if no precipitation occurs while it is high, no new oxygen is introduced to the system. To account for this scenario, two variables were created.

First, dissolved oxygen of the river water was multiplied by daily precipitation. The justification of this variable is that it provides a means of measuring total oxygen introduced into the system. Next, a three-day cumulative value of this variable was created using the following equation:

$$\boldsymbol{3\ day\ summed\ feature\ for\ row\ i} = \sum_{j=1}^{3} \boldsymbol{F_{i-j}} \qquad (5)$$

Preliminary modeling shows that wind direction is relatively important, but that wind speed may not be significant. Intuitively, we know that these two variables are related and that the combination of the two variables gives us more information than each variable separately, e.g. a 10 mph wind from the west may affect methane gas concentration more than a 1 mph wind from the west. These variables were combined using the following equation:

$$\boldsymbol{Wind\ feature} = \boldsymbol{Wind\ Speed} * \sin\left(\frac{2\pi[\boldsymbol{Wind\ Direction} - \mathbf{55}]}{\mathbf{360}}\right) \quad (6)$$

Equation 1 shows that the changes in gaseous concentration over time depends on gas velocity, gas dispersivity, and the molecular gas diffusion coefficient in the soil gas phase. Solving this equation does not fit within the scope of a data-driven approach to methane gas modeling; however, it does provide insights into useful variables. The molecular gas diffusion coefficient in the soil gas phase is proportional to T^1.5/p, where T represents temperature of the system and p represents pressure of the gaseous components of the system. This value was computed in the dataset by dividing recorded temperature, raised to the 3/2 by barometric pressure.

The final variables added to the dataset are lagged performance variables. Preliminary modeling attempts showed that it is not sufficient to simply use weather and water composition variables to predict methane concentration. By adding in lagged variables, we gain more insight into the activite components of the system. For this study, the closest lagged variables were four hours.

Initial Analysis

This project has two major components:

- Develop a regression model that accurately predicts the methane concentration of the landfill gas.
- Develop a classification model to predict if methane concentration will fall below 40% (1 = methane concentration <= 40%, 0 = methane concentration > 40%)

To develop a regression model for methane concentration, we must account for the drastic impact that extracting gas from the landfill has on methane concentration. Figure 2 illustrates this relationship. When the header valve is open, meaning that LFG is being extracted, we observe methane concentration of the LFG fluctuates drastically over the course of a day. When the header valve is closed, we observe methane concentration generally increases with time with much smaller fluctuations in the general path. Although this presents some challenges with the regression model, it opens another path of study.

By creating two regression models, one for each header valve scenario, we can develop a further understanding of the system. The regression model focused on the case when the header valve is closed has the potential to establish a baseline methane concentration growth rate. This baseline prediction may also have relevance in the regression model, focusing on the scenario in which the header valve is open and gas is being extracted.

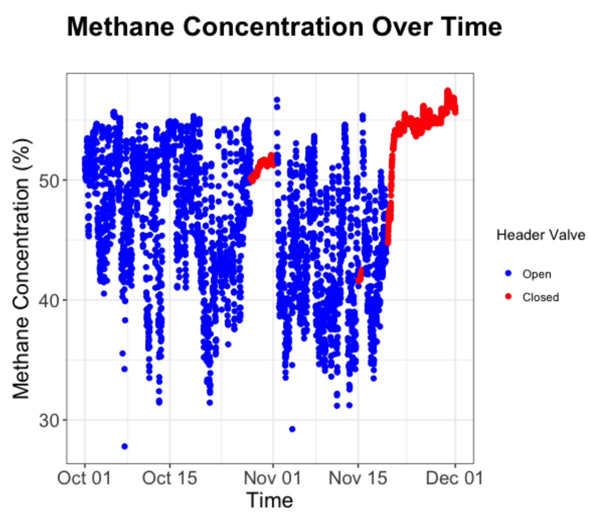


Figure 7. Displays LFG methane concentration at Rockingham County Landfill from 1 Oct 2012 to 31 Nov 2012.

According to the design specifications of the collection system engine used by the landfill, the engine is designed to utilize gas composed of 50% methane by volume. As concentration decreases, more stress is placed on the engine. Once methane concentration falls below 40%, the engine possesses a high risk of failure. Although the engine needs 40% methane concentration, the flare can likely operate with lower concentrations. Figure 3 shows the counts of each scenario for the binary variable created for the classification model.

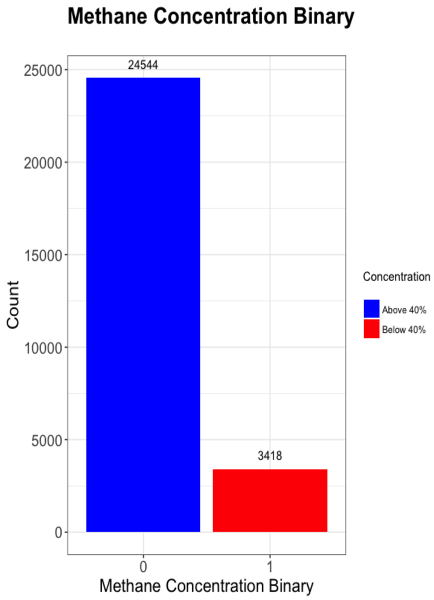


Figure 8. Displays the count of each scenario regarding LFG methane concentration.

We were unable to find any statistically significant relationship between the weather variables and methane concentration.

Model Construction
Prior to using any machine learning algorithm to develop a model, a baseline prediction was made. For regression models, this involved simply assuming the average methane concentration for every point in time. For classification, this involved assuming the methane concentration was always greater than or equal to 40%. This assumption was made due to the imbalanced structure of the dataset. This analysis was done in order to assess the effectiveness of each machine learning algorithm.

Multilinear regression and logistic regression were then used to provide a first attempt at the model. Both models utilize the same underlying math; however, logistic regression is used to predict the probability of an event occurring (methane concentration falling below 40%) while multilinear regression predicts a value (methane concentration). This is represented by Equation 7.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i \; for \; i = 1, 2, \ldots n \quad (7)$$

The final model utilized in the study is known as a multilayer machine learning algorithm. This technique utilizes a variety of algorithms and passes predictions from those algorithms to one or multiple higher-level algorithms. If multiple higher-level algorithms are used, the predictions from each algorithm are typically blended together to create an ensemble.

All models were developed using Python 3.6. On each model, a 4-fold cross validation was used.

4. RESULTS

Model Overview
To assess each classification model, sensitivity was used as the evaluation metric. Sensitivity is a measure of the model's ability to predict the positive case. This metric was selected because we want the model to detect an event which does not happen frequently. The regression model was assessed using mean absolute error (Equation 8). Mean absolute error was chosen because it shows, on average, how far off the estimate is from the true value. Table 1 shows the mean absolute error and sensitivity for each model.

$$Mean\ Absolute\ Error = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \quad (8)$$

	Model 0 Simple Average	Model 1 Multilinear/ Logistic Regression	Model 2 Multilayer Machine Learning
Mean Absolute Error	5.6%	3.1%	1.3%
Sensitivity	0%	0%	81.3%

Table 1. Mean absolute error and classification sensitivity for each of the three models used in this study.

Model 2 - Multilayered Machine Learning
Model 2, which was a multilayered machine learning model, clearly outperformed all other models used in this study. The basic structure of this model was as follows:

- The base level contained a combination of Multilinear/Logistic regression, Random Forests, Extremely Randomized Trees, Gradient Boosted Trees, Extreme Gradient Boosted Trees, K-Nearest Neighbors, and Neural Networks
- The intermediate level utilized two Gradient Boosted Tree structures from the LightGBM and XGBoost libraries.
- The final level was a weighted average between the two intermediate levels.

The mean absolute error of the predicted methane concentration and the actual methane concentration was 1.30% with a sensitivity of 81.3% (see Table 1).

The model had a classification accuracy of 95.0% (see Table 2). R^2=0.9032 when comparing the predicted methane concentration to actual methane concentration (see Figure 4).

	Predict Above 40%	Predict Below 40%
Actual Above 40%	23,935	808
Actual Below 40%	601	2,606

Table 2. Confusion matrix for classification model 2.

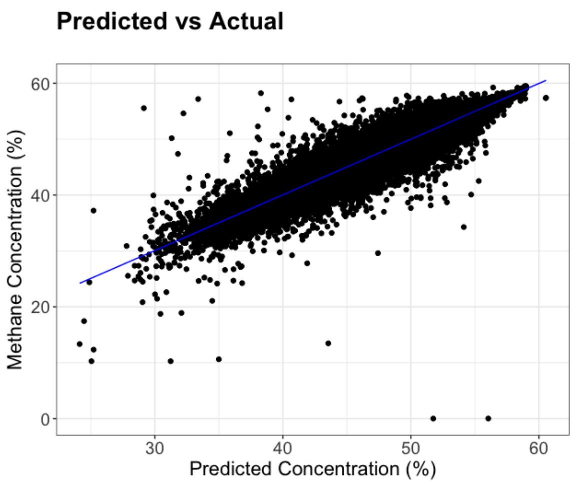


Figure 9. Predicted methane concentration of the LFG vs the actual methane concentration. R^2=0.9032.

5. DISCUSSION

The purpose of this project is to serve as an aid in the decision-making process for landfill operations. Our results show that through adequate data collection, variable manipulation, and proper predictive modeling techniques, LFG methane concentration can be predicted within a time horizon useful for operations.

By incorporating data science techniques into the day-to-day operations of a landfill gas collection system, the value and quantity of gas extracted from the landfill for generation purposes can be maximized. This predictive capability offers the potential for development of strategies guiding intervention and alteration of collection system operations prior to methane levels dropping below minimal useful thresholds, thus ensuring higher reliability and quality from landfill gas systems.

The classification model laid out in this paper has a sensitivity of 81.3%. If properly implemented, this model could prevent four out of five stress events from occurring. This result is quite impactful, as preventing engine stress and, in turn, preventing engine damage, can improve revenue and decrease the environmental impact of methane gas released by the landfill. This model also provides use to a fully automated system, as it can feed inputs into the system, changing its course of action.

With the ability to accurately predict engine stress events four hours in advance of the event occurring, landfill operators can alter their actions to alleviate the stress event. With the collection system database growing every day, we can improve the decision-making process, and modeling of events, by fully utilizing this data.

6. CONCLUSIONS

Landfill gas is a significant contributor to climate change, but its capture and use both reduces greenhouse gas emissions and provides a low-cost source of local renewable energy that can stimulate local economies. With the potential to benefit the local community, environment, and economy, the decomposition of solid waste should be seen as a potential resource. In fact, it should only be considered waste if it is not used.

This research provides a foundation for a new data-driven LFG model. By expanding data traditionally recorded during operations with secondary data that can serve as an appropriate proxy, it is possible to construct sub-hourly models of LFG flows using analytics techniques.

Additionally, this paper shows clearly how data analytics, combined with domain knowledge, can model a current relationship, predict a future one and eventually automate the process to transform industry.

Future research should investigate additional modeling techniques. Regression techniques were suboptimal in this case; however the Multi-layered machine learning model results suggest that Random Forests, Extreme Randomization, Gradient Boosting, Clustering techniques, and Neural Networks may improve predictions.

By moving forward with this research, we can assist waste management organizations with automation and better utilize data to maximize their resources through the incorporation of predictive analytics, ultimately converting more trash to treasure, while reducing the environmental impact of methane escaping from landfills.

7. REFERENCES

Abushammala, M. F., Qazi, W. A., Azam, M.-H., Mehmood, U. A., Al-Mufragi, G. A., & Alrawahi, N.-A. (2016). Economic and environmental benefits of landfill gas utilisation in Oman. *Waste Management & Research*, *34*(8), 717–723. https://doi.org/10.1177/0734242X16628983

Aghdam, E. F., Scheutz, C., & Kjeldsen, P. (2019). Impact of meteorological parameters on extracted landfill gas composition and flow. Waste Management, 87, 905-914.

Alexander, M. (1971). Microbial ecology. Wiley, New York.

Arigala, S. G., Tsotsis, T. T., Webster, I. A., Yortsos, Y. C., & Kattapuram, J. J. (1995). Gas generation, transport, and extraction in landfills. Journal of environmental engineering, 121(1), 33-44.

Buivid, M. G., Wise, D. L., Blanchet, M. J., Remedios, E. C., Jenkins, B. M., Boyd, W. F., & Pacey, J. G. (1981). Fuel gas enhancement by controlled landfilling of municipal solid waste. Resources and Conservation, 6(1), 3-20.

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. MIS Quarterly: Management Information Systems, 36(4), 1165–1188.

Chen, Y. C., Chen, K. S., & Wu, C. H. (2003). Numerical simulation of gas flow around a passive vent in a sanitary landfill. Journal of Hazardous Materials, 100(1-3), 39-52.

Farquhar, G. J., & Rovers, F. A. (1973). Gas production during refuse decomposition. Water, Air, and Soil Pollution, 2(4), 483–495. https://doi.org/10.1007/BF00585092

Garg, A., Achari, G., & Joshi, R. C. (2006). A model to estimate the methane generation rate constant in sanitary landfills using fuzzy synthetic evaluation. *Waste Management & Research*, *24*(4), 363–375. https://doi.org/10.1177/0734242X06065189

Harrill, D. J. (2014). Analysis of Changes in Landfill Gas OutPut and the Economic Potential for Development of a Landfill Gas Control Prototype (Master of Science, Appalachian State University). Retrieved from https://libres.uncg.edu/ir/asu/f/Harrill,%20David_2014_Thesis.pdf

Hashemi, M., Kavak, H. I., Tsotsis, T. T., & Sahimi, M. (2002). Computer simulation of gas generation and transport in landfills--I: quasi-steady-state condition. Chemical Engineering Science, 57(13), 2475-2501.

Hoornweg, D., & Bhada-Tata, P. (2012). What a Waste: A Global Review of Solid Waste Management (Review No. Knowledge papers no. 15). Retrieved from World Bank website: https://openknowledge.worldbank.org/handle/10986/17388

IPCC. (2017). IPCC — Intergovernmental Panel on Climate Change. Retrieved August 19, 2017, from IPCC — Intergovernmental Panel on Climate Change website: https://www.ipcc.ch/

Johari, A., Ahmed, S. I., Hashim, H., Alkali, H., & Ramli, M. (2012). Economic and environmental benefits of landfill gas from municipal solid waste in Malaysia. Renewable and Sustainable Energy Reviews, 16(5), 2907–2912. https://doi.org/10.1016/j.rser.2012.02.005

Kinman, R. N., Nutini, D. L., Walsh, J. J., Vogt, W. G., Stamm, J., & Rickabaugh, J. (1987). Gas enhancement techniques in landfill simulators. Waste management & research, 5(1), 13-25.

Lu, A. H., & Kunz, C. O. (1981). Gas-flow model to determine methane production at sanitary landfills. Environmental science & technology, 15(4), 436-440.

Monitoring of Operations (2000). 40 C.F.R. §60.756

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., … Zhang, H.

(2013). Anthropogenic and Natural Radiative Forcing. In: Climate Change 2013: The Physical Science Basis. In Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Retrieved from https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_Chapter08_FINAL.pdf

OECD. (2003). OECD Environmental Indicators: Development, Measurement, and Use [Reference]. Retrieved from Organization for Economic Cooperation Development website: http://www.oecd.org/environment/indicators-modelling-outlooks/24993546.pdf

Ozkaya, B., Demir, A., & Bilgili, M. S. (2007). Neural network prediction model for the methane fraction in biogas from field-scale landfill bioreactors. Environmental Modelling & Software, 22(6), 815-822.

Rachor, I. M., Gebert, J., Gröngröft, A., & Pfeiffer, E.-M. (2013). Variability of methane emissions from an old landfill over different time-scales. European Journal of Soil Science, 64(1), 16–26. https://doi.org/10.1111/ejss.12004

Reinhart, D. R., & Basel Al-Yousfi, A. (1996). The impact of leachate recirculation on municipal solid waste landfill operating characteristics. Waste Management & Research, 14(4), 337-346.

Reinhart, D. R., McCreanor, P. T., & Townsend, T. (2002). The bioreactor landfill: Its status and future. Waste Management & Research, 20(2), 172-186.

Sanchez, R., Hashemi, M., Tsotsis, T. T., & Sahimi, M. (2006). Computer simulation of gas generation and transport in landfills II: Dynamic conditions. Chemical Engineering Science, 61(14), 4750-4761.

Shrestha, S. O. B., & Narayanan, G. (2008). Landfill gas with hydrogen addition – A fuel for SI engines. Fuel, 87(17), 3616–3626. https://doi.org/10.1016/j.fuel.2008.06.019

Spokas, K., Bogner, J., Chanton, J. P., Morcet, M., Aran, C., Graff, C., … Hebe, I. (2006). Methane mass balance at three landfill sites: What is the efficiency of capture by gas collection systems? Waste Management, 26(5), 516–525. https://doi.org/10.1016/j.wasman.2005.07.021

Toerien, D. F., & Hattingh, W. H. J. (1969). Anaerobic digestion I. The microbiology of anaerobic digestion. Water Research, 3(6), 385–416. https://doi.org/10.1016/0043-1354(69)90002-5

Townsend, T. G. (2018). Landfill bioreactor design & operation. Routledge.

US EPA. (2016). Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2015 (Reports and Assessments No. EPA 430-P-17-001; p. 633). Retrieved from U.S. Environmental Protection Agency website: https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks-1990-2015

USACE. (2013). Landfill Gas Collection and Treatment Systems (Manual No. EM 200-1-22). Retrieved from US Army Corps of Engineers website: https://www.publications.usace.army.mil/Portals/76/Publications/EngineerManuals/EM_200-1-22.pdf

Warith, M. A. (2003). Solid waste management: new trends in landfill design. Emirates Journal for Engineering Research, 8(1), 61-70.

World Development Report 1992: Development and Environment, Andrew Steet. 1992. Oxford University Press, New York, NY. 308 pages. ISBN: 0-19-520876-5. (1994). Bulletin of Science, Technology & Society, 14(2), 117–118. https://doi.org/10.1177/027046769401400236

Xi, Y., & Xiong, H. (2013). Numerical simulation of landfill gas pressure distribution in landfills. Waste Management & Research, 31(11), 1140-1147.

Young, A. (1990). Volumetric changes in landfill gas flux in response to variations in atmospheric pressure. Waste Management & Research, 8(5), 379-385.

Young, A. (1992). The effects of fluctuations in atmospheric pressure on landfill gas migration and composition. Water, Air, and Soil Pollution, 64(3-4), 601-616