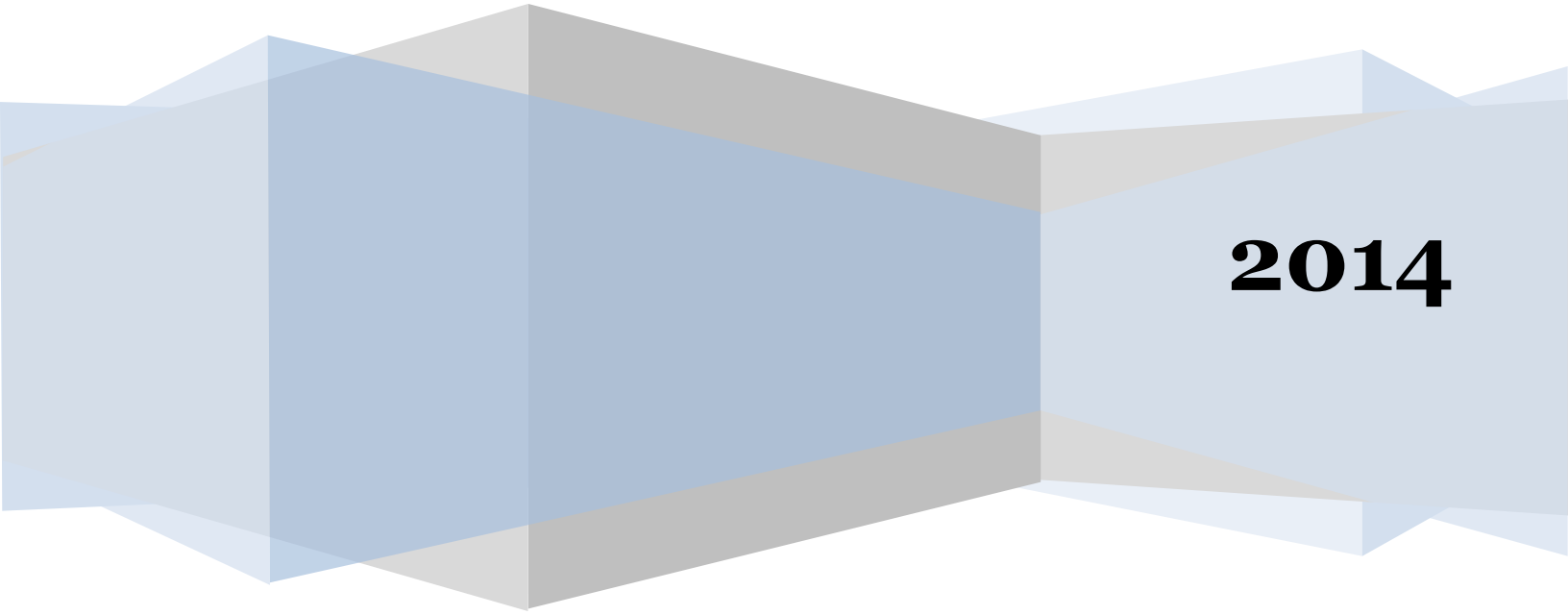
Tennessee Board of Regents
Maxine Smith Fellows Program

Greater than the Sum of Its Parts:

How to Effectively Use System-Level Data Warehousing in Higher Education

Melissa Irvin, Ed.D.
Tennessee Tech University
2014 Maxine Smith Fellow

Placement: Tennessee Board of Regents Central Office
Mentor: Chancellor John Morgan



2014

Greater than the Sum of Its Parts:
How to Effectively Use System-Level Data Warehousing in Higher Education

Table of Contents

Greater than the Sum of Its Parts:
How to Effectively Use System-Level Data Warehousing in Higher Education

EXECUTIVE SUMMARY

When the Complete College of Tennessee (CCTA) Act passed in 2010, the Tennessee Board of Regents (TBR) was challenged to uncover innovative methods to improve graduation rates and the number of Tennesseans with postsecondary education. During that same period, the Board approved the creation of a system-wide data warehouse to facilitate the state's shift to outcomes-based accountability and data-driven decision-making (J. Morgan, memorandum, May 30, 2014). The storage of system-level data from all institutions in TBR was entitled the "Common Data Repository and Electronic Data Warehouse (CDR/EDW)" and signaled an ambitious project to use the existing licenses for operational data stores (ODS) at the campus-level to create a timely, accurate tool for performing analytics on the system's "big data" (J. Morgan, memorandum, May 30, 2014).

The value of data warehousing is encapsulated in the movement away from simply storing records of daily business transactions; instead it is a systematic method for organizing both internal and external information that allows for easier retrieval and evaluate of the data (Weber & Weber, 2000). These types of data stores are increasingly valuable because their structure facilitates data mining and analytics. Data mining is the process of acquiring knowledge by analyzing patterns in the data to inform the decision-making process (Luan, 2002). Analytics takes things a few steps further. Instead of only looking for patterns or relationships in the data, analytics seeks to reveal trends and behaviors in order to predict what to expect in the coming months and even years (Wagner, 2011).

Despite the emergence of data warehousing and analytics as key strategies for higher education over the next few years (The New Media Consortium, 2013), there are still many challenges, particularly given the context of higher education, with implementing well-designed CDRs and creating the infrastructure to fully utilize them. Some of the common difficulties include:

- Lack of training and/or awareness of information technology staff on data warehousing and the resources necessary for analytics;
- Lack of communication and conflicting opinions between units and institutions about how data should be used and who should have access to it;

- Misconceptions that "big data", i.e., extremely large data sets, are notoriously inaccurate and difficult to use as well as maintain;
- Concerns in academia about overreliance on data to drive decisions in higher education;
- Navigating the different institutional terminology and language to form a standardized list of concepts; and
- Addressing issues of privacy to protect against misuse while still ensure the data is accessible to those who need it.

After reviewing case studies from a number of institutions (Cabrillo College, Georgia State University and Virginia Commonwealth University) as well as two data warehousing projects that combined student information from multiple institutions (the RAKETTI project in Finland and the Predictive Analytics Reporting [PAR] project), a set of strategies emerged to address anticipated challenges as the TBR's CDR becomes fully operational. Included recommendations for potential questions that could drive some of the analytics the system performs in the future.

Below is a brief summary of some suggested approaches to avoid or diminish the impact of any difficulties that may arise during the continuing implementation of the CDR:

- "Plant the seed that data is to be used, not just protected" (Wagner et al., 2012);
- Be sure to have precise definitions for common data concepts (i.e., low-income, veteran, etc.) available before users began retrieval and analyses in the warehouse (Wagner et al., 2012);
- New technology can be overwhelming and/or intimidated to some users (Kellen et al., 2013), so create avenues for brainstorming and constructive criticism;
- Avoid overreliance on the system support analysts from the technology vendors to explain how to navigate the warehouse and how to use any analytics tool (Kellen et al., 2013);
- Design an iterative process (Kellen et al., 2013) that allows users to be involved in the continual improvement of the data warehouse; and
- Harness word-of-mouth to promote the widespread use of the warehouse and to cultivate buy-in among end users (Kellen et al., 2013).

In addition, the following recommendations outline innovative ways to use a system-level data set to answer questions that would be time-consuming or unwieldy at the institutional level:

- Compare high-enrollment courses across institutions to discover if there are any differences in DFW rates;
- Evaluate how effective community college students transition to four-year institutions (see table on page 19 from 2012-2013 Fact Book);
- Compare patterns of course withdrawals across four-year institutions to see if differences in policy affect the number and/or timing of withdrawals;
- Evaluate academic programs with high enrollment but low degree production to investigate when students leave the program; and
- Compare trends in student enrollment between institutions with a lower percentage FTE enrollment and those with higher percentage of FTE enrollment (see table on page 20 from 2012-2013 Fact Book).

The field of data warehousing and analytics will likely continue to grow in prominence within higher education over the next few years. It is important that institutions and their governing bodies are able to demonstrate the return on the investment from both the fiscal and human resources needed to set up these data systems. The Tennessee Board of Regents is uniquely positioned with the state's increasing focus on outcomes and the newly available access to system-level data to uncover areas of improvement that lead towards the type of gains that make the field of analytics so enticing.

INTRODUCTION

In recent years, the use of data to understand human behavior has permeated business, industry and even the political arena. The field of predictive analytics has moved rapidly from supporting marketing and other business practices to informing political campaigns or designing targeted outreach to encourage the use of social services (Luan, 2002; Weber & Weber, 2000). With the increasing focus on completion goals and other outcomes measures, higher education as a public institution has begun to gravitate towards the use of data to better assess why students are not enrolling or completing postsecondary education. Traditionally, data storage in higher education focused on simply recording daily processes in student services and providing summary reports to administrators as well as state and federal entities (Weber & Weber, 2000). Now individual institutions are investing in complex systems, such as Customer Relationship Management (CRM) programs, to correspond, track and later evaluate interactions with students, parents and even alumni.

However, as institutions have realized the power of using the years of student data to better inform policy and practice, Institutional Research (IR) offices are seeing increasing requests from academic administrators and student services officers for data to assist them in examining patterns of student choice and behavior (Lumina Foundation, 2005). As a result, the demand for more accessible, nimble and user-driven interfaces for data systems has increased; this approach to data storage and access allows every day end-users with basic technical knowledge to retrieve information without lengthy requests to IR or complex code from information technology staff (Weber & Weber, 2000). In the state of Tennessee, the passage of the Complete College Tennessee Act (CCTA) in 2010 has acted as a catalyst to change how stakeholders view progress and success in the state. Instead of only evaluating individual institutions, their goals, struggles and achievements, the state acknowledged the importance of having some consistent state-level goals for postsecondary education: including better transfer pathways between 2 and 4-year institutions, clearer mission differentiation with lesson program redundancy and a fully-realized system for community colleges as well as the colleges of applied technology.

With this shifting focus, the Tennessee Board of Regents' system office recognized the pressing need to have access to the type of student data previously only available at the institutional level in order to propose and assess their system-wide goals. In order for policy makers to make data-driven decisions and effectively assess progress towards goals, they needed timely access to data delivered in a format that allows that information to be understood and used. In June of 2010, the Tennessee Board of Regents (TBR) approved the design of a common data repository

(CDR) at the central office to include all four-year institutions, community colleges and colleges of applied technology within the system (J. Morgan, memorandum, May 30, 2014). This white paper will briefly examine the beginnings of CDR systems focusing on their purpose within the field of higher education and including a discussion on the challenges in implementing and using CDRs. In addition, several case studies will be examined to explore how institutions have used CDRs and predictive analytics to improve outcomes at their institution. Finally, the 2013-2014 system-wide data from the Tennessee Higher Education Commission (THEC) will be examined to uncover areas where TBR can apply best practices from individual institutions to analytics done at the system level.

THE BASICS OF DATA MINING & ANALYTICS

Traditionally data storage systems of the past were designed to record descriptive information and document processes (Weber & Weber, 2000): what happened, who did it happen to and when did it happen? As business practices in the sales and marketing industries began to discover the power in using this same information to drive decision-making, higher education quickly began to follow a similar path—first with recruitment and development and now into areas of student success and satisfaction (Luan, 2000). Instead of limiting the information to storage only, data warehousing uses a clear framework for internal and external data that allows stakeholders at any level to easily access and evaluate the data to interpret its value (Weber & Weber, 2000). Shifting from storing data to placing value on information to inform practice is essential in facilitating the adoption of more data-driven decision-making practices into higher education. Most two-year and four-year institutions have collected servers full of information on students for years, including pre-enrollment data and academic performance records but have only recently realized the wealth of knowledge gained by looking at this data in depth.

Weber and Weber (2000) indicate there are three primary benefits from setting up data warehousing: standardized data, user-friendly, and ease of use. When organizations go through the process of standardizing common data repositories (also referred to as electronic data warehouses or operational data storage), consistent methods and language are created for documentation that supports sharing and comparison. More user-friendly data systems move away from the use of technical language, so end-users are more comfortable working with data. Finally improving the ease of use makes it more likely that key stakeholders will embrace using data in their daily practice, especially if the process of retrieving the information is less complex and time-consuming. These criteria are particularly salient for higher education where data is often stored in multiple modalities and often accessed only for reporting purposes only. When

these criteria are not included, data is under-utilized, misinterpreted and inaccessible to those who need it the most.

The Power of Analytics

Simply finding effective ways to store information is not enough to continue the transition to a culture of data-driven decision-making. Data mining or analytics should be considered a close partner to the development of any CDR. Luan (2002) defines data-mining as "the purpose of uncovering hidden trends and patterns and making accuracy based predictions through higher levels of analytical sophistication…producing new observations from existing observations" (p.19). Once patterns begin to take shape, the information can be used to positively affect the type of decisions made and the responsiveness of the system. When analytics are connected with CDRs, the data available is more reflective of real-time needs at the organization. It also allows users to perform analyses with larger data sets (i.e., "big data") than ever before.

In the 2013 Higher Education edition of the NMC Horizon report, learning analytics is classified as an emerging technology on the mid-term horizon. Their research revealed that many higher education institutions would move towards embracing analytics in the next two to three years (The New Media Consortium, 2013). Using analytics can allow colleges and universities with growing student populations to develop personalized and supportive infrastructures for students such as (New Media Consortium, 2013):

- Designing more effective pedagogical methods and instructional software;
- Uncovering at-risk populations that need additional resources;
- Assessing whether interventions have been successful or need adjustments; or
- Creating adaptive learning environments to mirror the learning differences of students.

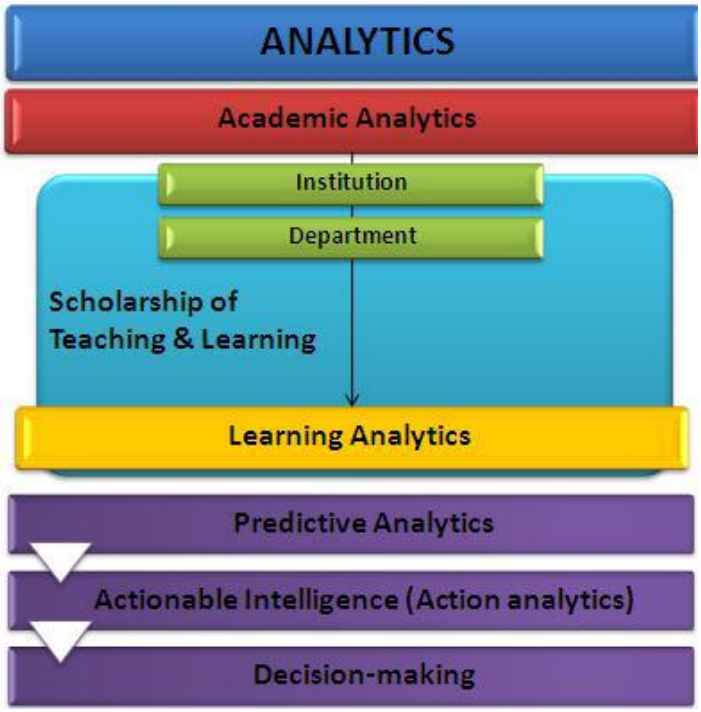
Postsecondary institutions that move purposefully towards incorporating analytics into their business practices will be on the forefront of modern-day higher education practice.

The conceptual framework proposed by authors working in the EDUCAUSE Learning Initiative (2012) seeks to further clarify how analytics works within higher education by identifying different fields within analytics (van Barneveld, Arnold, & Campbell, 2012):

- ***Analytics:*** The process of data-driven decision making at all levels;
- ***Academic Analytics***: Using data to inform the operational and financial decisions postsecondary institutions make;

- ***Learning Analytics***: Analysis of instructional and curricular information as well as any additional support resources to better reach established learning outcomes; and
- ***Predictive Analytics:*** Using large data sets to uncover patterns that can be effective in predicting behavior and/or decisions.

The figure below (van Barneveld, et al., 2012, pg. &) provides a visual depiction of the interrelated nature of these forms of analytics in higher education:



It is the inclusion of the student learning experience, i.e. learning analytics, into the field of analytics that makes it uniquely relevant to higher education (van Barneveld et al., 2012). Not only is this branch of analytics important for assessing institutional-level goals but it can drill down into issues at the department or even student level to better understand the nuances that influence the student educational experience.

THE CHALLENGES TO DATA WAREHOUSING & ANALYTICS

As with other innovations, using data warehouses and analytics is often met with obstacles that impede how quickly and effectively these resources are institutionalized into the culture of higher education. In the Horizon report, the New Media Consortium (2013) acknowledged that many higher education institutions have constraints within their organizational structure that may inhibit growth or the implementation of new technologies. Because analytics is still an

emerging field, IT database administrators and others may not yet have the necessary skills to design and implement a CDR or analytics programs (Picciano, n.d.). It is also essential that the data stored is always accurate and timely (Picciano, n.d.) to retain its value for decision-making. Offices and departments remain siloed from one another in ways that prevent effective data warehousing (Wagner, 2011). In addition, often times the offices collecting and using data have conflicting goals and ideas about what is an appropriate use of student information (Kellen at al., 2013; Wagner, 2011). Wagner (2011) also noted that most institutions do not have cultures where methods are in place to acknowledge patterns in information in ways that allow proactive responses.

"Lack of shared semantics" is a particularly difficult and complicated roadblock during the process of creating a centralized data warehouse for multiple institutions (Alonen, 2014). Managing the multiple ways that institutions can define concepts (such as transfer credit, concurrent degrees, prior term GPA, and academic status) can change what is measured and how a concept is measured and also constraint the ability to compare data from different institutions—especially when a common concept is measured and stored in different ways (Wagner et al., 2012). Those creating a common data repository must also recognized that most at the institutional level may be reluctant to share data (what type and how much) and allow others to access previously internal data (Wagner et al., 2012; Wagner, 2011). This fear may be linked to competition between institutions but also to concerns about data security and confidentiality laws. Postsecondary institutions must continue to be cognizant of privacy issues and protect data from misuse and abuse when attempting to broaden access (Picciano, n.d.).

Academia can be particularly distrustful of analytics; as an emerging field, there are still many areas that lack cohesive definitions and there is not yet an agreed on upon standard for assessing quality within the field (van Barneveld et al., 2012). Many make the assumption that larger data sets are inherently unwieldy to navigate and cannot be evaluated both quickly and accurately (Kellen et al., 2013). This may cause some to be hesitant to attempt to use or trust information found from the large data sets harvested from student information systems. In addition, many disciplines within academia will require new and different ways of approaching pedagogy which some seasoned academics will resist (van Barneveld et al., 2012). Finally, one of the common characteristics of the scholarship of teaching and research is the ability to share discoveries; this practice may clash with the business model of analytics, where the emphasis on gaining a competitive edge in the marketplace can override an institution's willingness to be transparent (van Barneveld et al., 2012).

CASE STUDIES ON USING CDRs AND ANALYTICS

Individual higher education institutions have stepped into using big data and predictive analytics in new and ever-expanding ways. In the past few years, trends have begun to emerge as institutions use student data to seek answers to many of the same questions (Kellen et al., 2013):

- Monitoring student uptake of campus services and activities like advising and tutoring;
- Uncover ways to target smaller groups of students for retention or graduation interventions;
- Provide immediate access to e-learning support tools based on real-time access to student performance data;
- Automated alerts when students fail to meet academic progress goals;
- Provide students with better recommendations of courses to take;
- Improved classroom scheduling to increase capacity without costly construction or renovations; and
- Forecasting how to best use different types of financial aid (merit and need-based) to maximize its impact on student success and provide an improved return on investment.

Nevertheless, what are effective ways of implementing these types of analyses? What type of student data is necessary to perform analytics like the ones mentioned above? How do institutions take the results and move towards actionable recommendations and positive change? By examining a few exemplary case studies, several strategies for best practices can be learned.

Cabrillo College

In 1998, a collection of postsecondary institutions consisting of one community college, Cabrillo College, and three California public universities agreed to create a data warehouse to track the success of transfer students (Luan, 2002). Their principle questions included: which students are more likely to transfer and which are likely to be successful after transfer to a university (Luan, 2002). Representatives from each institution collaborated to first determine the time frame of data to extract for the database and to standardize the different coding used to define similar information across years and institutions (Luan, 2002). Administrators eventually were forced to reduce the volume of data (such as condensing course type versus individual course performance information) to a manageable size and better quality (Luan, 2002).

From this larger combined data set, these institutions were able to develop a model producing a list of factors based on their importance: number of liberal arts courses taken, high school origin, race, planned work hours and others (Luan, 2002). Using this list of factors, the author recommends taking the patterns from the above model and developing a process for scoring currently enrolled students based on how they align with the key factors from the model (Luan, 2002). In addition, student support services can create targeted and personalized outreach to students tailored towards their score (Luan, 2002).

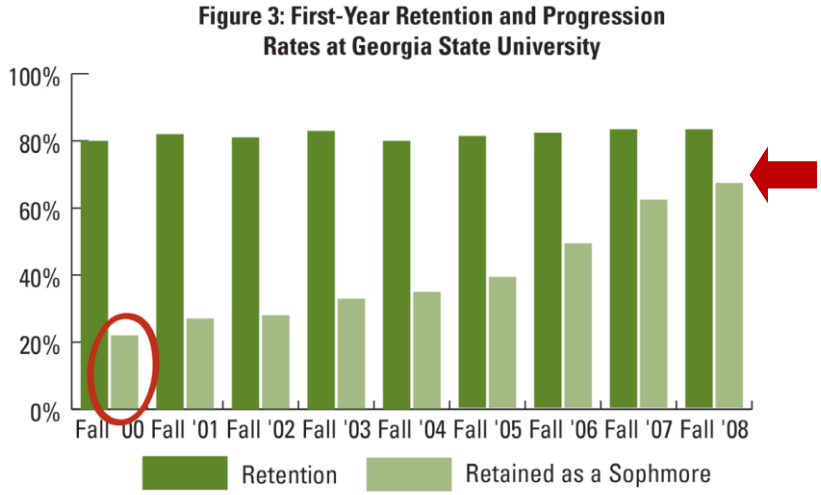
Georgia State University

At Georgia State University, administrators noticed that first-to-second year freshmen retention rates were robust (averaging 80-83%), but six-year graduation rates were considerably lower for the same cohort of students: around 41 percent (Education Trust, 2014). After performing some preliminary analyses on their freshmen, they discovered only 22 percent of students returning for their second year had earned enough credits during their first year to transition into sophomore standing (Education Trust, 2014). The institution then set out to perform some deeper analytics to discover what was contributing to the students' inability to earn enough credits early in their academic careers and positively impact completion (Education Trust, 2014). What they discovered was the students who only earned on average ten credits hours during their first semester often did not pass or withdrew from key milestone courses; these students also limited course loads to manage costs (Education Trust, 2014).

From these results, faculty and administrators at Georgia State developed a handful of student interventions (Education Trust, 2014):

- Offering learning communities for almost three-quarters of the freshmen population to increase the number of hours earned in their first semester;
- Redesigned key milestone courses with high DFW rates by including supplemental instruction and hybrid instructional methods; and
- Providing at-risk students with the opportunity to earn additional credits during their first summer semester (e.g., Summer Success Academy).

These strategies have helped Georgia State see dramatic gains in the number of students starting their second year of college as sophomores (22 to 67% in eight years – see graph on the following page) and improve their six-year graduation rate by six percent (41% to 47%) over the last five years (Education Trust, 2014).



Virginia Commonwealth University

One of the challenges of using extremely large data sets to perform analytics is knowing where to begin to look for the patterns that could provide insight. At Virginia Commonwealth University, a team tracked a cohort of freshmen for two years to examine everything from demographic data, pre-enrollment data, income status, living arrangements and academic records for each semester (Education Trust, 2014). Overall 72 percent of freshmen remained enrolled at VCU by year three, but the university found revealing patterns within the students who did remain at the institution: "academic work (high school and college GPA) was more important than measured academic ability (standardized test scores)" in positively affecting retention (Education Trust, 2014, pg. 11). Academic standing was also revealed to be an important factor in determining whether a student would be retained (Education Trust, 2014).

VCU also examined course withdrawal patterns to see if there was any correlation to four-year, five-year and six-year graduation rates. Data revealed that students with no more than one dropped class on average graduated on time in four years while students who graduated in six years typically had about eight dropped courses (Education Trust, 2014). Administrators were also able to see that international students were withdrawing from much larger numbers of courses (approximately 25%) compared to domestic students who withdrew from an average of 7 percent of attempted courses (Education Trust, 2014).

From their analyses, Virginia Commonwealth made changes their course withdrawal policy including considering the last date of withdrawal and limiting the number of withdrawals (Education Trust, 2014). The administration also agreed that focusing resources on improving second and third year retention rates would positively affect completion rates (Education Trust,

2014). Subsequently the university has seen improvements in first-to-second year retention rates as well as an increase to the number of students in good academic standing after year one from 73 percent to 82 percent using strategies such as: extended summer orientation programs, a year-long required cohort program for freshmen and intrusive student support services (Education Trust, 2014).

RAKETTI project in Finland

In 2008, the Finnish Ministry of Education and Culture, along with the Finnish postsecondary institutions, undertook a project to create a place to share academic data from all institutions; the RAKETTI project was born (Alonen, 2014; Remes, 2010). RAKETTI is an abbreviation for the Finnish phrase "RAkenteellisen KEhittämisen Tukena TIetohallinto" which translates in English to "information services as part of the universities' structural reform" (Remes, 2010). The catalyst for this project was a lack of uniformity in key concepts institutions used to guide their practices (Remes, 2010) and the financial implications of facilitating long-term improvements at the institutional level:

> *"We offer centralized and common service for 50 universities and polytechnics in Finland. The estimated cost of building a data warehouse for one higher education institution is EUR 1.2 million, so if each one were to create its own warehouse, the total cost would be 60 million…By creating a shared data warehouse, solutions made once can easily be copied. This can reduce the costs per institution to less than a tenth. At the same time, the information will all be in the same form and suitable for comparison due to the common concept model" (Frantti, 2009).*

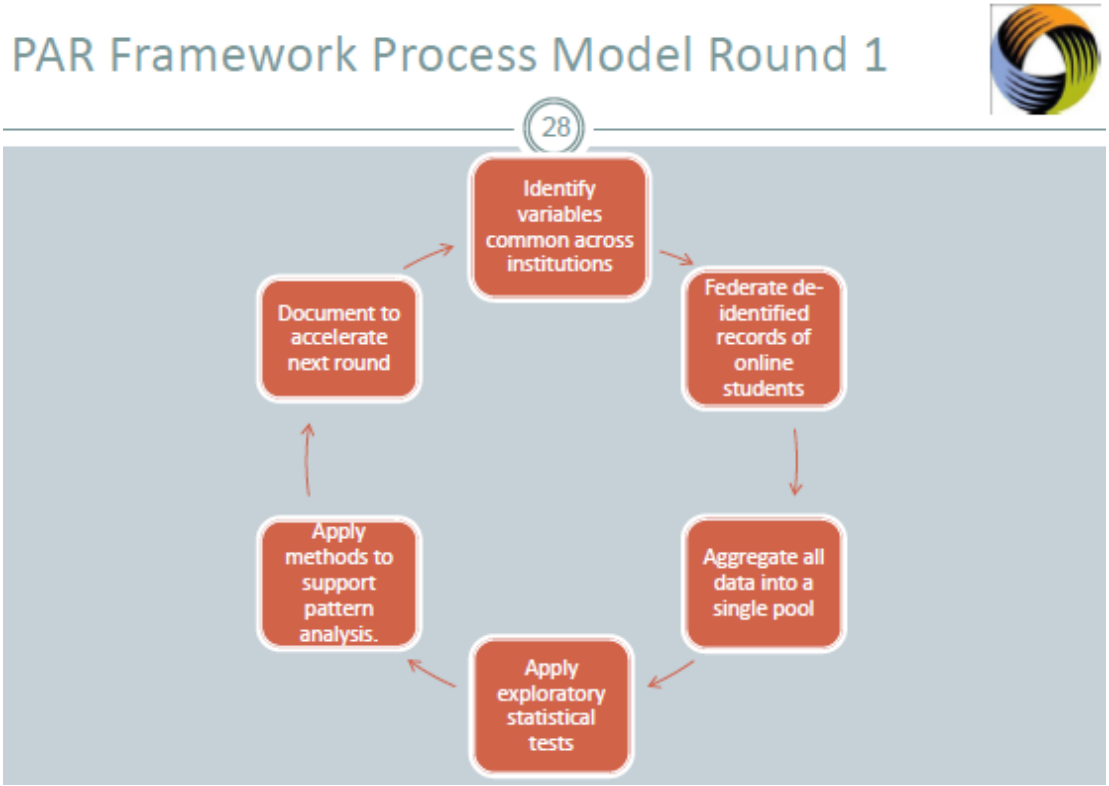
This coalition envisioned the project as four separate initiatives to standardize administrative record keeping, evaluate academic services related to both instruction and support services, create an academic database to search faculty scholarship and compare business practices between institutions (Remes, 2010).

The implementation of this project has been lead by the participants' desire to create some type of common framework for storing, managing, retrieving, analyzing and comparing their data. Starting with a group of pilot institutions, the first stage of establishing the data warehouse was creating common terms for concepts used by all institutions, which resulted in a glossary of over 300 words (Frantti, 2009). After groups for a variety of departments at the participating institutions agreed upon the concept definitions, the pilot programs began testing the accuracy and reliability of the collected data concerning student, financial, human resources and facility

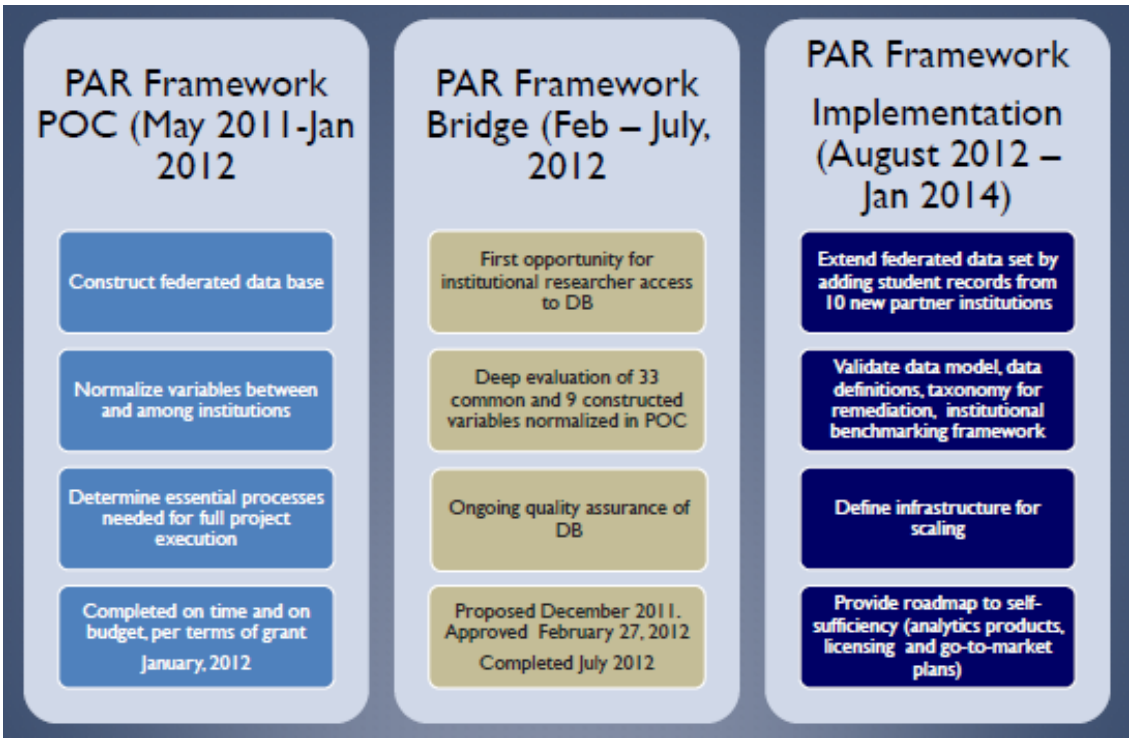
administration (Frantti, 2009). In addition, the pilot tested the most effective methods to transfer data into the warehouse and produce internal and external reports (Frantti, 2009); this stage of the implementation process continues to be one of the considerable challenges in the EDW's design (Alonen, 2014). As the project designers move into the next stage of this warehouse's evolution, their focus shifts towards developing methods to improve the effectiveness and efficiency of how data is transferred between systems (Alonen, 2014).

Predictive Analytics Reporting (PAR) Framework (http://www.parframework.org/)

The WICHE Cooperative for Educational Technologies (WCET) received a grant from the Bill and Melinda Gates Foundation early in 2011 to combine data from six pilot institutions into one data warehouse to study attrition, academic progress and completion (Wagner, 2011). This was a considerable challenge since it involved creating a CDR with over 640,000 student records from six institutional partners from across the country: two four-year institutions (University of Hawaii System, University of Illinois—Springfield), two community colleges (Colorado Community College System, Rio Salado College), and two for-profit institutions (American Public University System, the University of Phoenix) (Wagner et al., 2012). The grantees developed a clear model to navigate the initial stage of the warehouse's creation (Wagner, 2011):



During the process of creating the Predictive Analytics Reporting (PAR) framework, the participating institutions uncovered a number of other problematic issues that arise when constructing massive data warehouses for data mining and analytics. Project designers realized the importance of investing considerable time at the beginning of the process to ensure the data from across institutions is comparable (Wagner et al., 2012). Participating institutions quickly recognized the value of having access to data stored in one place and the power of collaboration across institutions (Wagner et al., 2012). They discovered that while most were reluctant to share much data at first that the institutions' shared interest in improving student outcomes and the project's success helped to break down barriers where the majority were eager to share information and to work with the data (Wagner et al., 2012). The PAR project continues to expand by adding more institutions and using the patterns uncovered during the pilot to establish benchmarks that participating institutions can use to set reasonable goals and better measure progress (Wagner et al., 2012):



RECOMMENDATIONS FOR PRACTICE

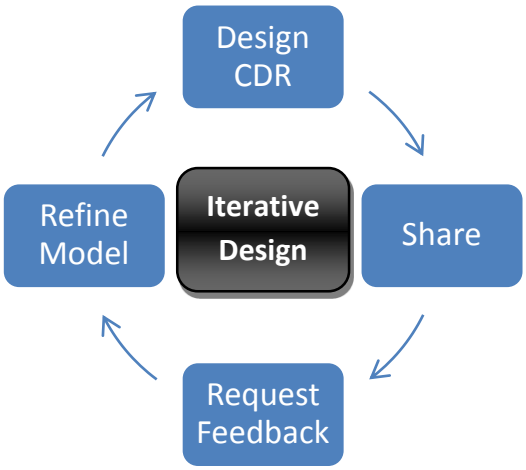
The Tennessee Board of Regents project to create a central data repository using existing ODS licenses is on track to allow access to retrieve data, run reports and perform analyses by October 2014 (J. Morgan, memorandum, May 30, 2014). As the implementation moves forward, it is

important to consider how some key lessons learned from other institutions can inform practices for TBR to better improve the adopt and use the system's CDR.

Navigating the Challenges of Implementing Data Warehousing

The following recommendations are designed to encourage the use of the CDR and to suggest methods to effectively navigate the obstacles that may arise as the warehouse is actively utilized.

- ***Addressing Concerns About Data Accuracy & Security***
  - "Plant the seed that data is to be used, not just protected." (Wagner et al., 2012)
  - Design an iterative process (Kellen et al., 2013) that allows users to be involved in the continual improvement of the data warehouse.



  - Emphasize early that there is no "finished product" when working with data and stakeholders should expect (and support) continual innovation and improvements to the system.
- ***Working with Those Hesitant and Unfamiliar***
  - Be sure to have precise definitions for common data concepts (i.e., low-income, veteran, etc.) available before users began retrieval and analyses in the warehouse (Wagner et al., 2012).
  - New technology can be overwhelming and/or intimidated to some users (Kellen et al., 2013), so create avenues for brainstorming and constructive criticism.
  - Harness positive word-of-mouth from early adopters to promote the widespread use of the warehouse and to cultivate buy-in (Kellen et al., 2013).
  - Create mechanisms for pilot users to share their successes with using the data warehouse with others to stimulate interest in how it can be beneficial for their department or campus.

    - Advance towards developing customizable reports and/or dashboards that allow institutions to see data relevant to their interests.
- ***Addressing Infrastructure & Resources***
    - Avoid overreliance on the system support analysts from the technology vendors to explain how to navigate the warehouse and how to use any analytics tool (Kellen et al., 2013).
    - Know the skill set of TBR's information technology team and have ambassadors available who are capable of "translating" the technical jargon into digestible guidance and support for end users.

Envisioning Analytics for TBR's Data Warehouse

Higher education institutions must consider this observation by Richard Alford when determining how to use the data they have collected: "Evidence never contains its own explanation" (Alford, 1998, p.29 as cited in Lumina Foundation, 2005). The Lumina Foundation encourages those at postsecondary education to shift how they choose to look at the use of data in decision-making by transitioning from "a culture of evidence" to "a culture of inquiry". In a "culture of evidence", institutions engage faculty and administrators at all levels, especially those in institutional research, in assessing program effectiveness and student outcomes as well as sharing those discoveries with the campus community (Lumina Foundation, 2005). When embracing a "culture of inquiry", the focus changes to consider who is interpreting and how data is interpreted instead of simply reporting results (Lumina Foundation, 2005). It is vital to consider why certain information is collected, who defines and interprets the data and how analyses are shared with others at the institution (Lumina Foundation, 2005). The authors of the Lumina report make the following recommendations for establishing a "culture of inquiry" (2005, pg. 2):

- Work to identify and address problems by purposefully analyzing data about student learning and progress;
- Engage in sustained professional development and dialogue about the barriers to student achievement; and
- Have the capacity for insightful questioning of evidence and informed interpretation of results.

As the Tennessee Board of Regents embarks upon utilizing its CDR for the entire system, a clear outline for guided inquiry can reflect the tenets listed above and hopefully demonstrate the

value of continuing to develop the capabilities of the centralized data warehouse for the system. To that end, below are some recommended projects derived from the institution-level best practices described earlier in this report and aligned with the system's college completion agenda that capitalize on the CDR's ability to combine or compare data from multiple institutions:

- ***Evaluate how effective community college students transition to four-year institutions (see table below from 2012-2013 Fact Book).***
    - Are there characteristics that make students more successful as transfer students from a community college to a four-year institution?
    - Are students from certain community colleges more likely to earn a bachelor's degree than others?
    - Do students from particular two-year institutions graduate with more excessive credits than those from comparative institutions? Do some four-year institutions have transfer students from two-year institutions graduating with more excessive credits than other institutions?
    - Compare the differences in graduation rates between students who start at a two-year institution and graduate from a four-year institution, students who start and graduate from a four-year institution and those that start at one four-year institution and graduate from a different four-year institutions.

Undergraduate Transfer Student Activity by Sending Institution into Receiving Public University Fall 2013

Sending Institution	Total Sent	APSU	ETSU	MTSU	TSU	TTU	UoM	UTC	UTK	UTM	UTHSC
Austin Peay State University	96		*	26	18	8	8	*	12	15	
East Tennessee State University	124	8		31	*	11	10	9	48	*	
Middle Tennessee State University	229	21	*		36	29	49	26	35	22	*
Tennessee State University	86	*	*	45		*	14	*	*	7	
Tennessee Technological University	121	*	8	48	7		7	20	19	*	*
University of Memphis	103	6		37	6	*		*	13	22	12
University of Tennessee, Chattanooga	221	21	9	67	9	10	23		67	15	
University of Tennessee, Knoxville	247	12	42	70	*	15	36	39		22	*
University of Tennessee, Martin	102	8	*	20	*	*	47	7	*		*
University Total	**1,329**	**87**	**73**	**344**	**88**	**81**	**194**	**117**	**207**	**111**	**27**
Chattanooga State Community College	369	6	38	38	6	29	*	239	11	*	
Cleveland State Community College	110		16	15	*	15	*	44	18		
Columbia State Community College	306	19	6	169	16	15	*	27	23	27	*
Dyersburg State Community College	154	6	*	8	*	*	51	*	*	78	*
Jackson State Community College	261	8	*	39	*	*	95	6	*	94	*
Motlow State Community College	353	*	*	218	20	81		14	15	*	
Nashville State Community College	450	65	*	148	152	45	*	12	13	8	*
Northeast State Community College	303	*	268	9		7	*		16		
Pellissippi State Community College	633	19	50	32	*	63	*	36	425	*	
Roane State Community College	299	*	45	13		143	*	12	83		
Southwest Tennessee Community College	605	6	6	31	26	*	457	*	11	13	49
Volunteer State Community College	413	61	11	133	82	66	*	17	35	*	
Walters State Community College	322	*	186	20		8	*	8	94		
TBR Community College Total	**4,578**	**201**	**633**	**873**	**313**	**479**	**622**	**421**	**753**	**229**	**54**

- ***Compare trends in student enrollment between institutions with a lower percentage FTE enrollment and those with higher percentage of FTE enrollment (see table below from 2012-2013 Fact Book).***
    - Are the student characteristics that cause students to enroll in fewer hours?
    - Are their institutional barrier (i.e., course availability, course scheduling, etc.) that limit the number of courses enroll in each semester?

FTE as a Percent of Headcount Enrollment Fall Terms 2003 and 2008 - 2013							
Institution	**2003**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**
TBR Community Colleges							
Chattanooga State Community College	63.9%	62.9%	63.5%	64.3%	63.5%	64.8%	63.1%
Cleveland State Community College	70.4%	65.8%	69.3%	69.0%	68.3%	68.2%	65.2%
Columbia State Community College	66.8%	64.6%	63.9%	63.4%	62.0%	62.2%	63.4%
Dyersburg State Community College	72.3%	63.3%	65.2%	64.4%	61.8%	61.7%	58.6%
Jackson State Community College	68.7%	64.0%	64.8%	63.4%	66.0%	63.3%	59.3%
Motlow State Community College	70.0%	65.8%	66.0%	63.5%	61.3%	61.2%	60.6%
Nashville State Community College	55.7%	55.9%	58.1%	56.7%	57.5%	57.5%	57.0%
Northeast State Community College	64.4%	65.9%	67.5%	68.2%	68.3%	66.5%	66.4%
Pellissippi State Community College	66.3%	65.0%	65.0%	65.1%	64.6%	66.1%	64.4%
Roane State Community College	70.1%	68.1%	67.3%	63.8%	60.9%	62.4%	62.3%
Southwest Tennessee Community College	63.3%	63.2%	65.0%	62.4%	63.0%	61.8%	62.5%
Volunteer State Community College	63.3%	63.3%	65.3%	64.0%	62.7%	62.0%	60.9%
Walters State Community College	65.4%	69.0%	69.7%	69.1%	68.0%	67.3%	65.3%
TBR Community College Total	**65.1%**	**64.0%**	**65.0%**	**64.0%**	**63.4%**	**63.2%**	**62.2%**
TBR Universities							
Austin Peay State University	82.4%	79.8%	74.3%	79.0%	78.2%	80.1%	80.5%
East Tennessee State University	85.5%	83.9%	84.0%	85.3%	85.2%	84.1%	83.9%
Middle Tennessee State University	86.2%	84.1%	83.6%	82.6%	81.8%	81.4%	81.6%
Tennessee State University	85.5%	81.1%	79.6%	79.7%	77.7%	77.7%	80.2%
Tennessee Technological University	82.5%	79.4%	83.5%	81.2%	81.1%	83.2%	87.5%
University of Memphis	79.0%	78.7%	78.4%	77.6%	77.5%	78.3%	77.4%
TBR University Total	**83.4%**	**81.4%**	**81.0%**	**81.0%**	**80.4%**	**80.8%**	**81.4%**
UT Universities							
University of Tennessee, Chattanooga	83.7%	86.1%	86.6%	91.3%	86.4%	85.7%	87.6%
University of Tennessee, Knoxville	90.1%	90.0%	90.5%	88.7%	86.2%	87.8%	88.3%
University of Tennessee, Martin	91.2%	83.2%	82.9%	82.1%	86.5%	87.2%	88.2%
University of Tennessee, Medical Health Sci Center	97.1%	100.0%	100.0%	97.0%	100.0%	100.0%	100.0%
University of Tennessee Total	**89.3%**	**88.7%**	**88.9%**	**88.6%**	**87.1%**	**87.9%**	**88.8%**
University Total	**85.4%**	**84.0%**	**83.8%**	**83.6%**	**82.7%**	**83.3%**	**84.0%**
Grand Total	**77.6%**	**76.5%**	**76.3%**	**75.6%**	**75.0%**	**75.4%**	**75.4%**

- ***Compare high-enrollment courses across institutions to discover if there are any differences in DFW rates.***
    - What are the differences in DFW rates in high-enrollment classes at two-year and four-year institutions?
    - Do institutions with higher rates of student success in higher enrollment classes have different delivery models for the course content?

- ***Compare patterns of course withdrawals across four-year institutions to see if differences in policy affect the number and/or timing of withdrawals.***
    - Which institutions have fewer students who graduate with an excessive number of course withdrawals?
    - Are their academic programs and/or student classification levels where students are more likely to accumulate course withdrawals?
    - Are there different patterns at different institutions of the timing of course withdrawals during a semester?
- ***Evaluate academic programs with high enrollment but low degree production to investigate when students leave a program.***
    - Are there particular academic programs with high enrollment who do not graduate a represent percentage of their population? Does this vary between institutions that offer similar programs (i.e., Nursing)?

CONCLUSIONS

After the passage of the Complete College Tennessee Act of 2010, the climate for postsecondary education in the state decidedly shifted from a focus on enrollment within individual institutional models to a system-wide set of strategies intended to improve student outcomes. The use of data warehousing and the field of analytics is an obvious tool that the Tennessee Board of Regents can use to plan, implement and assess these goals. The challenge lies in the system's ability to take a relatively new field and adapt it to meet the needs of the sixth-largest higher education system in the country. Although the implementation stage is a difficult part of making the TBR's central data repository a reality, the real test begins when the stakeholders attempt to use the CDR, at both the system-level and at the institutional level, to evaluate progress or make recommendations for change. Who will be the early adopters? How can the system encourage involvement at the institutions? Will the system be responsive in addressing early issues with accuracy, efficiency, and accessibility? TBR must be willing to support the continual development and maintenance of the CDR and its analytical process to best ensure that it becomes an institutionalized part of decision-making in Tennessee higher education.

REFERENCES

Alonen, M. (2014, September 30). Semantic interoperability architecture for higher education and research. Retrieved October 4, 2014 from https://confluence.csc.fi/pages/viewpage.action?pageId=42798717.

The Education Trust. (January 2014). *Learning from high-performing and fast-gaining institutions*. Washington, D.C.: Yeado, J., Haycock, K., Johnstone, R., & Chaplot, P. Retrieved January 17, 2014, from http://www.edtrust.org/higher_ed_practice_guide.

Frantti, A. (2009, April 1). CSC—IT Center for Science Ltd. *A shared data warehouse for higher education institutions*. Retrieved June 2, 2014, from http://www.csc.fi/english/csc/publications/cscnews/2009/4/data_warehouse.

Kellen, V., Recktenwald, A., and Burr, S. (2013). *Applying big data in higher education: A case study*. Arlington, MA: The Cutter Consortium.

Luan, J. (2002). *Data mining and its applications in higher education*. New Directions for Institutional Research, 113, 17-36.

Lumina Foundation for Education. (December 2005). *Data don't drive: Building a practitioner-driven culture of inquiry to assess community college performance*. Boston, MA: Dowd, A.C.

Morgan, J. (2014, May 30). *TBR CDR/EDW* [memorandum]. Nashville, TN: Tennessee Board of Regents.

The New Media Consortium. (2013). *NMC Horizon Report: 2013 Higher Education Edition*. Austin, Texas: Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., Freeman, A., and Ludgate, H.

Picciano, A. (n.d.). *The evolution of big data and learning analytics in American higher education*. Journal of Asynchronous Learning Networks, 16(3), 9-20.

Remes, M. (2010, February). CSC—IT Center for Science Ltd. *RAKETTI on its way to better information*. Retrieved June 2, 2014, from http://www.csc.fi/english/csc/publications/cscnews/2010/2/raketti.

Tennessee Higher Education Commission. (2014). 2013-2014 Tennessee Higher Education Fact Book. Retrieved from http://www.tn.gov/thec/Legislative/Reports/2014/2013-2014%20FACTBOOK.pdf.

van Barneveld, A., Arnold, K., & Campbell, J. (2012). *Analytics in higher education: Establishing a common language*. Washington, D.C.: EDUCAUSE.

Wagner, E. (2011, July). *Data changes everything*. Presentation given at the Sloan C 4th annual international symposium, San Jose, CA. Retrieved from http://wcet.wiche.edu/par/presentations.

Wagner, E., Iboshi, P., Sharkey, M., & Sherrill, J. (2012, November). *When learning analytics meet big data: The predictive analytics reporting (PAR) framework*. Presentation given at EDUCAUSE conference, Denver, CO. Retrieved from http://wcet.wiche.edu/par/presentations.

Weber, R., & Weber, J. (2000). The use and value of data warehousing in higher education. *Mountain Plains Journal of Business and Economics*, *1(1)*. Retrieved June 2, 2014, from http://www.mountainplains.org/articles/2000/general/mpa15.html.