

2013

## 2013 Big Data Conference Program

Big Data Conference

Follow this and additional works at: [https://scholarworks.gvsu.edu/bigdata\\_conference2013](https://scholarworks.gvsu.edu/bigdata_conference2013)

---

### ScholarWorks Citation

Big Data Conference, "2013 Big Data Conference Program" (2013). *2013 Presentations*. 23.  
[https://scholarworks.gvsu.edu/bigdata\\_conference2013/23](https://scholarworks.gvsu.edu/bigdata_conference2013/23)

This Article is brought to you for free and open access by the Big Data Conference at ScholarWorks@GVSU. It has been accepted for inclusion in 2013 Presentations by an authorized administrator of ScholarWorks@GVSU. For more information, please contact [scholarworks@gvsu.edu](mailto:scholarworks@gvsu.edu).

## GVSU “Big Data” Conference 2013 April 26, 2013

### **Max Eckard, Kyle Felker, and Carlos Rodriquez (University Libraries)**

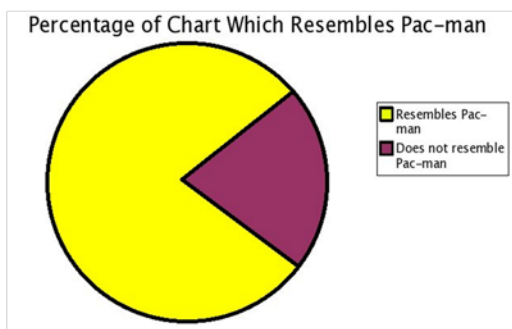
The Library & Big Data: What Role Should the Library Play?

*Academic Libraries have historically been responsible for collecting, describing, providing access and preserving the output generated by and for the scholarly, academic and research community. Many of the information management and organization skills, expertise and knowledge required to provide these services can be applied to university’s rapidly growing digital data and information resources. This session will provide participants an opportunity to share their Big Data needs, requirements, and challenges and begin to explore some possible new roles and responsibilities to support this growing need to manage and preserve the university’s “Big Data.”*

### **Paula Lancaster (Special Education, Foundations, and Technology Department, College of Education)**

From Gut Reaction to Graphic Representation: Using Data to Make Educational Decisions

*Increasingly educators in the PK-12 schools are required to use data collected from various sources to guide their instructional and placement decisions and to share these data with parents. Questions have arisen about the extent to which “Big Data” can be harnessed to inform decisions across a district, building, and classroom.*



### **8:30 Poster Session**

#### **Daniel Hodges (Graduate Student – Department of Statistics)**

Applications of High Performance Statistical Computing  
*Using a hybrid approach of statistical modeling and high performance computing enables statistical models to be built from large data sets. This approach enabled the modeling of atmospheric conditions on Lake Michigan in order to gain a better understanding of surface wind dynamics across open water. With an understanding of surface wind dynamics more accurate forecasting models can predict wind speeds in stable conditions.*

#### **Nancy Richard (University Archivist)**

Special Collections & University Archives – From 15th Century Printing to Big Data: Challenges of Collection, Preservation, and Access into the Future

*Learn about the various challenges brought about by inclusion of digital records in GVSU’s Special Collections and University Archives and the various opportunities to address these challenges facing faculty, administration and the Department of Special Collections & University Archives.*

#### **Max Eckard (Metadata & Digital Curation Librarian)**

Data Management Planning for the “Lone Wolf” Researcher  
*Since January 18, 2011, the NSF has required a supplementary document of no more than two pages describing a Data Management Plan for proposed research in all proposals. This poster details the components of a data management plan, and should be helpful for faculty members who are applying for NSF, NIH or other federal grants, as well as forward-thinking, “lone wolf” faculty who simply want to manage and describe their data in such a way that it is suitable for long-term preservation and for sharing with researchers at other institutions and across disciplines.*

**Paul Stephenson and Marco Benedetti (Dept. of Statistics)**

Charting Procedures for Monitoring Michigan Geospatial Data

*This poster will demonstrate charting procedures that can be used to monitor Michigan spatial data over time. In addition, this poster will present an examination of the power associated with these monitoring procedures by simulating the effect of an increase in one, two or three regions in the state of Michigan.*

**9:00 Presentation Session #1****9:00 Welcome – Maria Cimitile (Provost’s Office) and Edward Aboufadel (Dept. of Mathematics)****9:15 Paul Stephenson (Dept. of Statistics)**

A Random Sample of Big Data Resources That Are Readily Available Via the Internet

*The rapid accumulation of data has the potential to change the way institutions, governments, and businesses make decisions. One challenge facing educators will be identifying sources of big data that could be infused into their classes. The author will survey a variety of resources (e.g., GapMinder) on the Internet that may help one locate and utilize big data for research or in class.*

**9:30 Carlos Rodriguez and Max Eckard (University Libraries)**

Thinking Long-Term: The Research Data Lifecycle beyond Data Collection, Analysis and Publishing

*Research data often has a longer lifespan than the research project that created it. Researchers may continue to work on data after funding has ceased, follow-up projects may analyze or add to the data, and data may be re-used by other researchers. In this presentation, we will discuss the iterative research data lifecycle, focusing on life after the creation and analysis of research data, and beyond research article publication.*

**David Zeitler (Department of Statistics)**

Capabilities of the Statistics Department Analytics Server

*The deluge of data available to researchers is increasing the likelihood that analysis will be beyond the capabilities of personal computers. To meet this increased need the statistics department has setup a mid-level analytics computer to meet these computationally intensive needs. This talk will briefly discuss the current and planned capabilities of this new resource.*

**Dan Frobish (Department of Statistics)**

Predicting Survival Probability Based on Gene Expression Levels

*Predicting survival probability (or any other response variable) from gene expression data presents some challenging problems, because the number of variables is typically much larger than the sample size. I will briefly mention a few methods that are used for dimension reduction.*

**12:45 “Birds of A Feather” Roundtables****Jerry Scripps (School of Computing and Information Systems)**

Data Analytics for Any Student

*At one time, algebra was a specialized skill, known only to a few. More recently, word processing was also something that only some employees knew. Over time some skills become important enough to introduce to the broader population. With the onset of Big Data, will Data Analytics become a necessary skill? If so, at what depth? I would like to have a discussion involving educators from many different disciplines about the future of this field.*

**Mary Schutten (Department of Movement Science)**

Big Data: Academic Achievement; Big Data: Fitness; Big Data: SES in school aged children

*The purpose of this study was to examine the association between physical fitness and academic achievement and to determine the influence of socioeconomic status (SES) on the association between fitness and academic achievement in school-aged youth. Compared to all other variables, SES appears to have the strongest association with academic achievement. However, high fitness levels are positively associated with academic achievement in school-aged youth.*

**Agnieszka Szarecka (Department of Cell and Molecular Biology)**

Simulations of Antibiotic Resistance: Mining the Trajectory Data  
*Computer simulations of enzyme dynamics reveal details of the mechanisms through which beta-lactamases select and inactivate the antibiotic molecules. We have a great need to design antibiotics that will withstand the emerging bacterial resistance and long time-scale data will help us with that challenge.*

**Debbie Lown (Department of Biomedical Sciences)**

Examining Disparities in Food Access and Enhancing the Food Security of Underserved Populations in Michigan (FAiM)

*We will conduct in-depth interviews with a stratified (by type of food outlet) sample of food retailers in 18 towns and cities in Michigan. Retailers will be found using Michigan Department of Human Services (MDHS) and Michigan Department of Agriculture databases. The 2010 Census will be used to analyze census block group (CBG) data to determine the proximity-weighted racial composition of neighborhoods as well as the level of segregation and poverty in neighborhoods. Strategies identified for enhancing food security of underserved populations in the state will be shared at state and local levels.*

**9:45 Greg Wolffe (School of Computing and Information Systems)**

What's Behind the Curtain?

*A look at the infrastructure of cloud computing and how it supports Big Data research.*

**10:00 Jerry Scripps (School of Computing and Information Systems)**

In Praise of Forgetfulness; Machine Learning for All

*Machine Learning is both highly complex and oddly familiar. The techniques are employed in many highly visible websites but are also oddly apparent in our everyday logic. This talk describes machine learning to a highly intelligent audience in an oddly intuitive way.*

**10:15 Jonathan Leidig (School of Computing and Information Systems)**

Very Large Digital Libraries

*Digital libraries are systems that store large quantities of diverse content. They are well suited to store, index, manage, and preserve the information produced and consumed by big data efforts.*

**10:30 Break****10:45 Presentation Session #2****10:45 David Zeitler, Mehmet Sözen, and Charlie Standridge (Department of Statistics and School of Engineering)**

Reflections on Lake Michigan Wind

*A laser wind sensor mounted on a floating buoy is being used to assess the wind energy resources in Lake Michigan. Each second, the laser wind sensor measures wind speed and direction at 6 heights ranging from 75m to 175m. Analyses compare wind speed and direction over the course of 7 months at various heights as well as estimating the power potential of the wind.*

**11:00 GVSU Web Team**

Educating the World through Data

*How do people learn? What knowledge are people searching for? How do we get them that information and in what form? As a team who has functioning websites with over 30,000 visits per months and a global reach, we implemented analytics in order to generate data to answer these questions. The use of analytics and data has transformed our sites and has been monumental in our learning as we take more steps in reaching our goal of open access to education across the globe.*

**11:15 Laura Kapitula (Department of Statistics)**

Can “Big Data” Tell the Whole Story about Teacher Quality?

*In this talk I will discuss the use of student test scores and value-added modeling in teacher evaluation systems. Using estimates from value-added models in high stakes decisions can have unintended consequences. Systems that use both “Big Data” and “little data” maybe better able to discriminate teacher quality than systems that rely on only student test scores for evaluation.*

**11:30 Andy Baalerud (Graduate Student – Disability Support Resources)**

Students with Disabilities: Promoting Academic Success

*This presentation will highlight data pertaining to the development and participant-satisfaction of the Learning Specialist Curriculum within Disability Support Resources. This initiative is a resource used to help students with disabilities develop effective learning strategies and study skills.*

**11:45 Paul Leidig, Paul Stephenson, Jonathan Leidig, Jerry Scripps, Guenter Tusch and David Zeitler (School of Computing and Information Systems and Dept. of Statistics)**

Curricular Initiative Regarding a Minor in Data Science

*Over the past several months faculty in statistics and computing & information systems have met to develop the framework for a new minor in Data Science. We will discuss our progress on this curricular initiative and elicit feedback.*

**12:15 Lunch & Lightning Round!**

**Neal Rogness (Department of Statistics)**

Quantitative and Qualitative Analysis of the myGVSU (Climate Study) Survey

*The myGVSU survey was administered in February 2011 to all GVSU faculty, staff, and students, of which over 7500 individuals participated. The survey contained 106 questions which translated into over 1000 unique variables since some questions had multiple parts; further, approximately 25 questions were open-ended and over 25,000 written comments were provided by respondents. The combination of quantitative and qualitative data provides a rich snapshot of how respondents felt about the overall climate of the university at the time of the survey’s administration.*

**Greg Wolffe (School of Computing and Information Systems)**

Data-mining for Computational Public Health Research

*Using information extracted from mobile datasets to improve public health prediction, planning and policy-making.*

**John Gabrosek (Department of Statistics)**

Breakdowns in Common Statistical and Graphical Techniques for Big Data

*Many common statistical techniques, such as hypothesis testing, break down in the presence of massive data sets. What common techniques need alternatives in the face of Big Data?*

**Carlos Rodriguez (Associate Dean of Technology & Information Services)**

Issues and Challenges: The Various V’s of Big Data

*Briefly share Big Data issues and challenges as they related to volume, variety, velocity, veracity, validity and volatility.*