

**TDWI** WEBINAR SERIES

# Enabling the Citizen Data Scientist: How to Build a Bigger Business Impact with Analytics

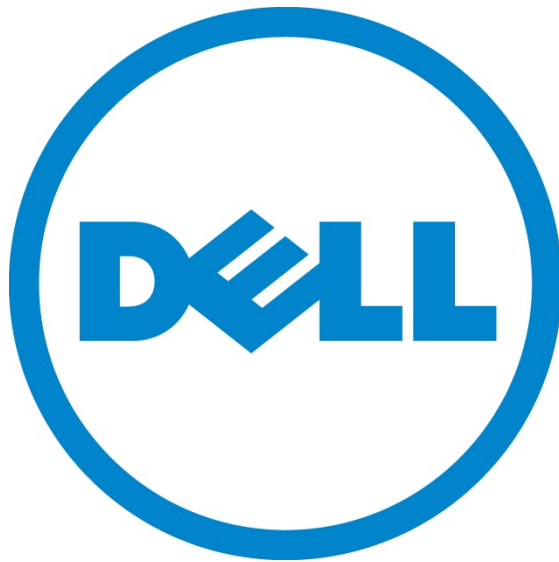
**David Stodder, Director of Research for Business Intelligence, TDWI**  
**Robert J. Lake, Data Scientist, Cisco Systems**

January 20, 2016

[tdwi.org](http://tdwi.org)

tdwi  
Advancing all things data.

# Sponsor



# Statistica

**TDWI** WEBINAR SERIES

**tdwi**  
Advancing all things data.

# Speakers



David Stodder  
Research Director,  
Business Intelligence  
TDWI



Robert John Lake  
Senior Manager,  
Data Science,  
Cisco

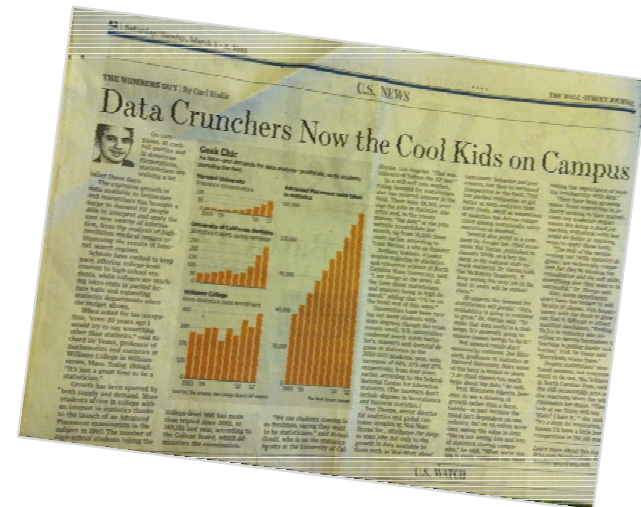
# Agenda



- Defining data science and its role in the analytics culture
- How data science is different, and complements BI and business analysis
- Data science success factors
- Empowering “citizen” data scientists
- Concluding best practices
- Robert J. Lake
- Q&A

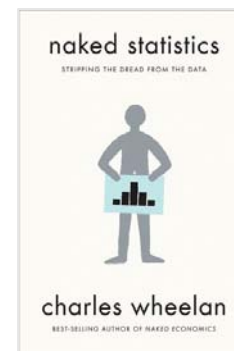
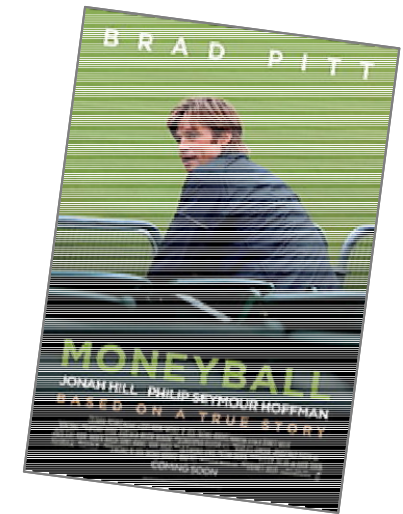
# Data Scientists: Cool, Sexy, and Rare

- Surge in analytics creates jobs for data scientists and software engineers who can crunch data and build algorithms
- “The next sexy job in next 10 years will be statisticians” – *Hal Varian, Google Chief Economist*
- “Unicorns”: Hard to find, hard to keep; data scientist shortage



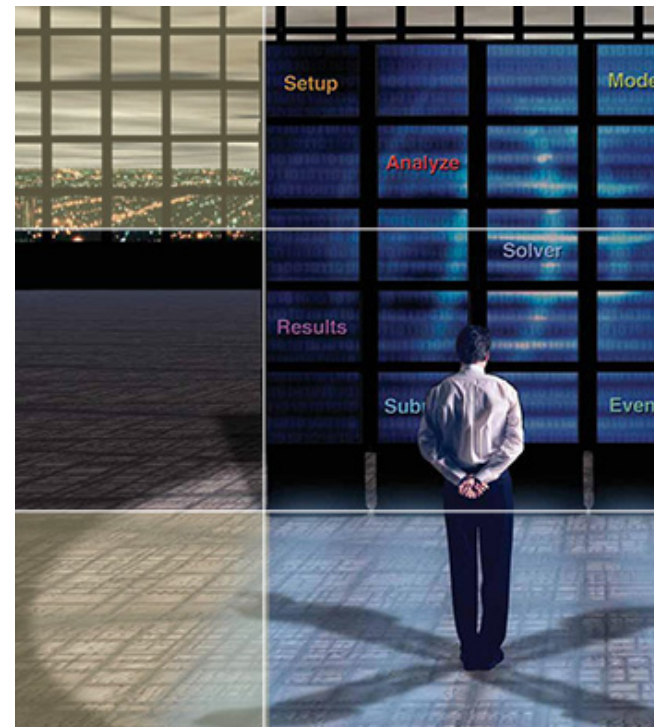
# Data Science Becomes Part of Culture

- Politics, movies, and bestsellers reflect strong interest in data science
- Chief data scientist, chief data officer: new c-level titles
- Automating decisions for speed and quality in the Age of Algorithms: “If a decision is algorithmic, it belongs in code.” – *Marc Demarest, Noumenal*



# Analytics: The Driving Force

- **Competitive advantage:** the quest to develop and deploy better analytics than your rivals
- **The unknowns:** Finding patterns that could affect outcomes
- **Data ROI:** “There’s so much data just sitting there doing nothing”



# Analytics: Sharpening Focus on Outcomes

- **Looking beyond reports:** What is the desired business outcome?
  - *Making room for experimentation and discovery with more varied data*
- **Optimization:** Using analytics to reduce waste and improve efficiency
- **Analyzing** real vs. forecast performance and expectations; identifying outliers and anomalies
- **Aspiring to** predictive and proactive engagement





# Not Just Insights: Value & Outcomes

- “You succeed with analytics when you stay focused on the end goal. It isn't enough to find patterns in the data and highlight trends and outliers in fancy charts, or deliver insights that can potentially drive business value. **Your analysts must actually create business value.**”
- “If nothing changes because of [analysts’] insights, then they haven't added any value to the business.”

– *Ken Rudin, Former Head of Analytics, Facebook and now at Google;  
from interview with W. Eckerson at TechTarget*

# What is Data Science?

- “The study of the generalizable extraction of knowledge from data” – *Vasant Dhar*
- Contributions from many fields: statistics, computer science, software engineering, data visualization, and more
- Often covers processes for acquiring and cleaning data (data preparation) to prepare datasets
- Data science: About applying scientific methods to explore and test hypotheses
  - *Many data scientists come from hard science fields*
- Continuous experimentation using hypothesis-free discovery
  - *Machine learning*



# What is a Data Scientist?

- Experts who can extract insights from data tsunamis
  - *R, Python, other tools/langs*
- Common job duties, drawn from position openings:
  - *Apply machine learning to optimize ad selection*
  - *Develop algorithms to take into account hundreds of different attributes to build user profiles*
  - *Interface with BI, product engineering, product mgmt*
- Data scientists tap rich datasets to transform insights into real-world products
  - *Be passionate about using data to drive strategy and product recommendations*
  - *Crunch the numbers one minute and communicate effectively to deliver complex, data-driven findings*
  - *Research new ways of modeling behavior*

Thanks to Datafloq for some of these points.

# BI vs. Analytics/Data Science: Differences

## BI/OLAP Realm

- Traditionally limited to query and reporting on narrow selection of structured data
- Reporting: Have to know ahead what you want to know; limits on interaction
- Can deliver precise answers: But are they relevant to decisions?
- Metadata, schema limits: No one best and consistent way to define, categorize all information; silos abound

## Analytics/Data Science

- Solving unknowns: Asking right questions, not just getting answers
- Investigative, iterative “what-if” data inquiry; questions leading to more questions, with different variables
- Multivariate analysis using multiple types of data
- Trends, relationships, correlations, across sources – structured relational, semi-structured, unstructured

# Comparing Business Analysts and Data Scientists

## Business Analysts

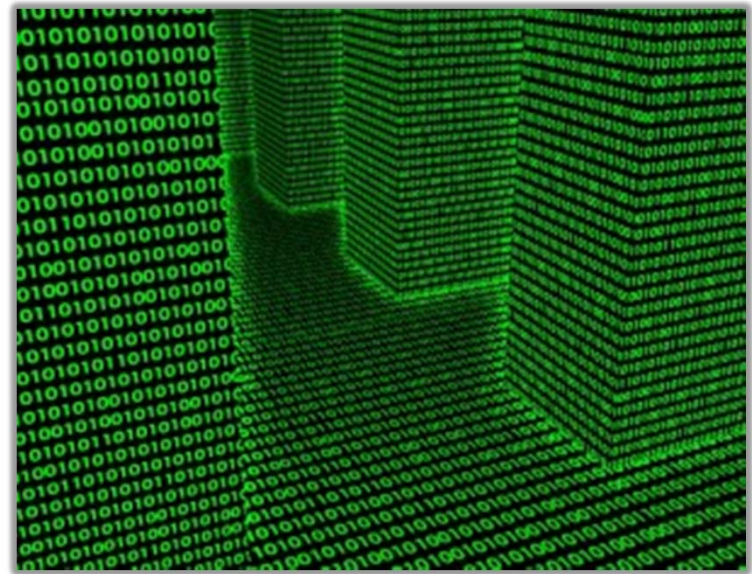
- Background in business; evaluate past, present, and future business performance
- Use analytical models and approaches
- Define business problems and translate statistical analysis into data-driven BI
- Research, interpret, and visualize raw data using predictive, prescriptive, and descriptive analysis

## Data Scientists

- Design, develop, and deploy algorithms through statistical programming; manage large amounts of data; create visualizations to aid in understanding
- Mine data and write machine learning algorithms
- Employ statistical languages and tools
- Create visualizations
- Design and structure datasets

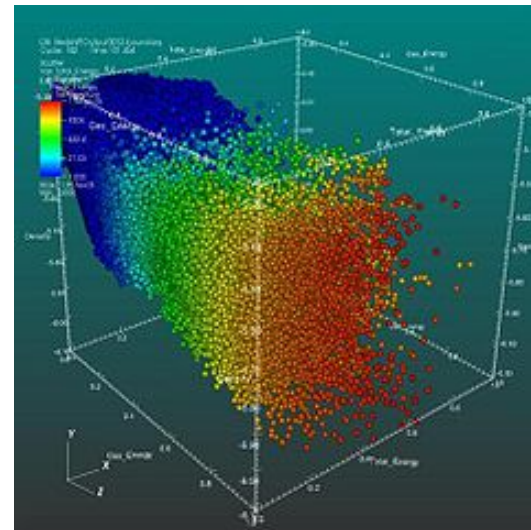
# Statistics and the Influence of Big Data

- Sample sizes explode: massive, fine-grained sources for analysis
- Focus more on correlations than causality
- From estimating basic probabilities to running complex simulations with large numbers of variables
- Variable selection critical to value and method of calculation



# Data Science Use Cases

- **Recommendation systems:** eBay, Netflix, Spotify, etc.
- **Customer intelligence:** Understanding and predicting user behavior
- **Marketing...and Elections:** Constructing accurate voter targeting models; increasing voter, donor participation
- **Healthcare:** Analyzing relationship between treatments and outcomes
- **Internet of Things:** Making sense of huge and fast data streams
  - *Vehicles, wearables, and situational awareness*



# Chasing Unicorns: A Better Approach?

- Creating an effective team for data science
  - *Drawing on business, IT, and app development*
  - *Empowering SMEs*
- Good communication skills: Not everyone has them
- **Needed:** Expertise in the business domain, analytics, statistics, data preparation, and more
- Many organizations seeking to train internal employees for data science

What is your organization's current experience with, and/or future plans for employing, "data scientists" to meet objectives for customer and social media data analytics?

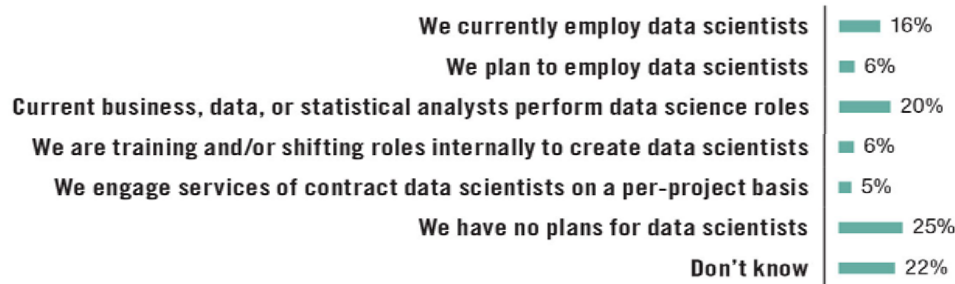


Figure 10. Based on answers from 423 respondents.

From "Customer Analytics in the Age of Social Media," TDWI Best Practices Report, Third Quarter 2012



# Communication Skills: Critical

- Universally seen as a key element of success with data science
  - *Challenging assumptions and conventional wisdom takes good communication*
  - *Speaking the business language; what are the key business questions to solve?*
- Data silos: Gaining access is often about gaining trust
  - *“What’s in it for me?”*

➤ **Best practice:** Identify business stakeholders and keep them involved

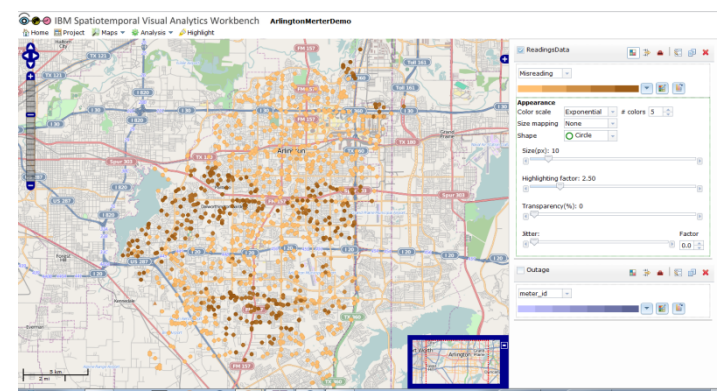
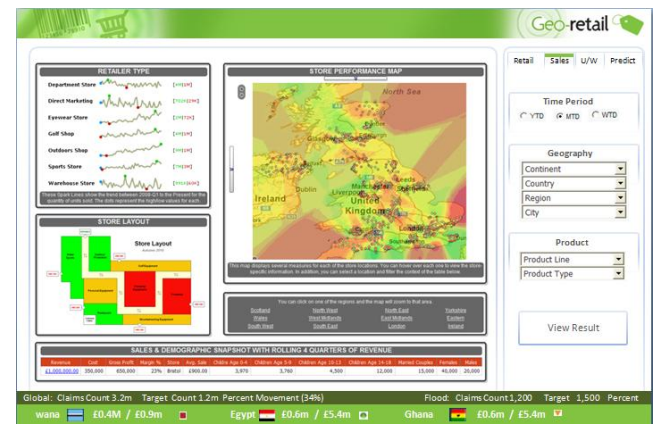


# Data Scientist Success Factors

- ✓ “Intellectual curiosity...to find meaning in the chaos”
- ✓ Ability to ask the right questions to drive business impact; “surfacing and testing assumptions about what people think drives behavior or business metrics”
- ✓ Can engage in iterative experiences with business users to ensure impact. Not easy: Overcoming difficulties in the first knowledge transfer meeting
- ✓ Can tell stories with data: from millions of data points, tell us what is the guidance?
  - ✓ *Powerful use of data visualization*

# Visualization: Enabling Citizen Scientists

- **Fusion:** Analytics, test-and-learn data exploration, and advanced computation matched with visualization
- **A visual path:** data interaction through filtering, comparing, and correlating visual data relationships
- **Business data laboratory:** Enabling exploration of who, what, when, why behind events and transactions



# Visualization: Key to Self-Service Analytics

- **Business-Driven Analytics:**  
“Citizen” subject matter experts want to more control of data access and analysis
  - *Visualization eases the path for nontechnical users*
  - *Trend toward self-service data preparation*
- **Satisfying the variety of users and requirements:**
  - *Visualization is key to personalization of the data and analytics experience*



# Operationalizing Analytics for ROI

- Business SMEs often want to deploy analytics within processes and operations asap
- Moving from descriptive analytics focused on what and why
  - *To “prescriptive,” recommending actions*
  - *Predictive modeling to forecast what might happen given variables*

- **Best practice:** Use technology and best practices to reduce time it takes to develop and deploy models
- Clean up analytic workflows; find repeatable patterns



# The Importance of Governance

- “Citizen” data science must not turn into the “wild west”
  - *IT confidence is key to accessing data*
- Sensitive data must be protected and secure
- Training is vital for citizen data scientists to realize value; can be championed through governance committees

- **Best practice:** Establish data stewardship that includes business and IT
- Use governance to help guide training and project leadership



# Wrapping Up: Best Practices

- Focus on delivering value, not just insights
- Understand what makes data science different
- Create effective teams rather than chasing unicorns
- Aim at how you can operationalize analytics

- Make communication a priority for data science teams
- Employ data visualization to improve communication and empower citizen data scientists
- Address governance and security

# Thank You!



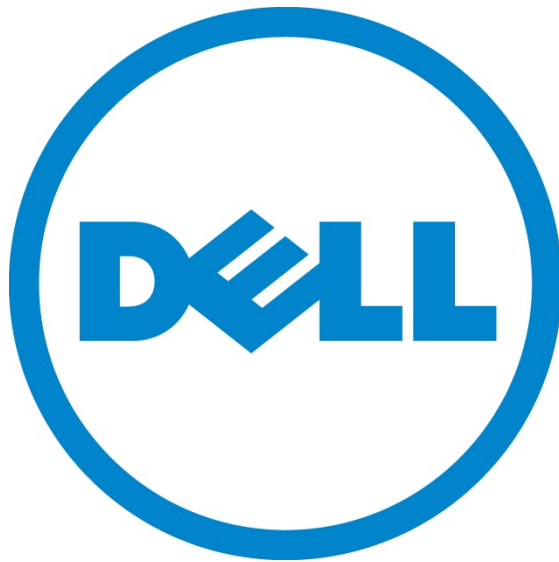
David Stodder  
Director of Research for Business Intelligence  
TDWI ([www.tdwi.org](http://www.tdwi.org))  
[dstodder@tdwi.org](mailto:dstodder@tdwi.org)  
(415) 859-9933



# Poll Question

- What is your biggest challenge in expanding data science and enabling “citizen” business users to do more with analytics?
  - *Data governance and security concerns*
  - *Data quality, and preparation challenges*
  - *Lack of adequate tools and technologies*
  - *Business users lack of training in analytics*
  - *Hard to identify business requirements*
  - *Other*

# Thank You to Our Sponsor



# Statistica



# Developing Community Data Scientists

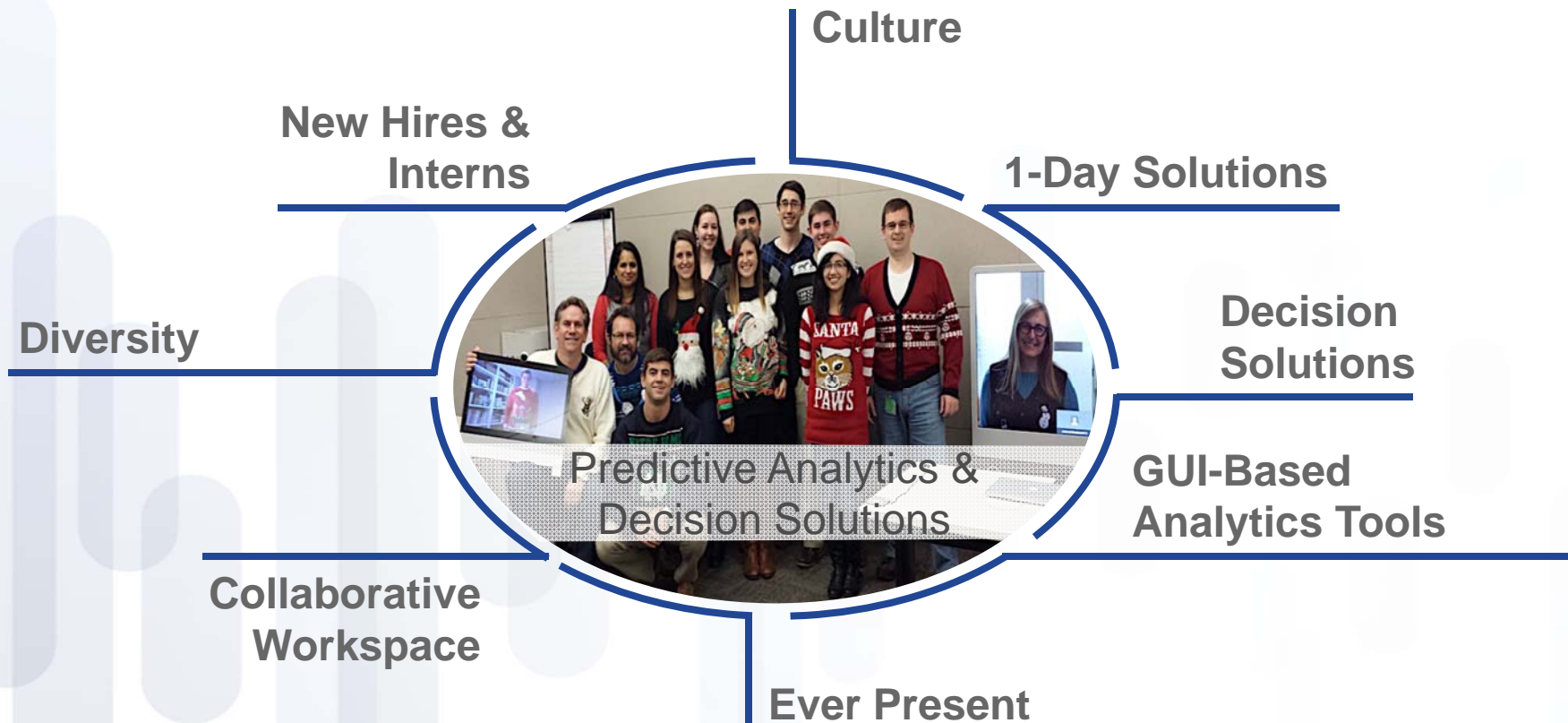
Robert John Lake  
Senior Manager, Data Science  
January 20, 2016



**Robert John Lake**  
Senior Manager, Data Science



# Developing Data Scientists





**CISCO**

*TOMORROW starts here.*

# Questions?



# Contact Information

- If you have further questions or comments:

David Stodder, TDWI  
dstodder@tdwi.org; @dbstodder

Robert J. Lake, Cisco  
rjl@cisco.com



# Learn More in Las Vegas 2016!

## TDWI Conference

“Advancing All Things Data”

Las Vegas, NV | January 31 – February 5, 2016

<http://www.tdwi.org/LV2016>

\*

## TDWI Executive Summit

“Make Analytics Pervasive in Your Organization”

Las Vegas, NV | February 1-2, 2016

<http://www.tdwi.org/LV2016/ES>