JUMPING INTO BIG DATA:

How the Media and Entertainment Industries are Getting Started

October 23, 2014

#CREbigdata

cre

council for research excellence

OPENING REMARKS

Richard Zackon, CRE
Ceril Shagrin, CRE Council Chair
Stacey Schulman, Big Data Chair

cre

cre
council for research excellence

ABOUT THE CRE

The Council for Research Excellence is a body of senior research professionals, formed in 2005 to identify important questions about audience measurement methodology and to find, through quality research, the answers to those questions.

The Council provides the Nielsen client community a means to undertake research projects no one company could undertake on its own.

cre

council for research excellence

CRE MEMBER COMPANIES

BIG DATA COMMITTEE MEMBERS

Stacey Schulman, Chair

- Michele Buslik
- Jon Cogan
- Laura Cowan
- Pete Doe
- Janice Finkel-Greene
- Sam Garfield
- Paul Hockenbury
- Annette Malave
- Michele Meyer
- Dan Murphy
- Rosemary Scott
- Ceril Shagrin
- Howard Shimmel

cre

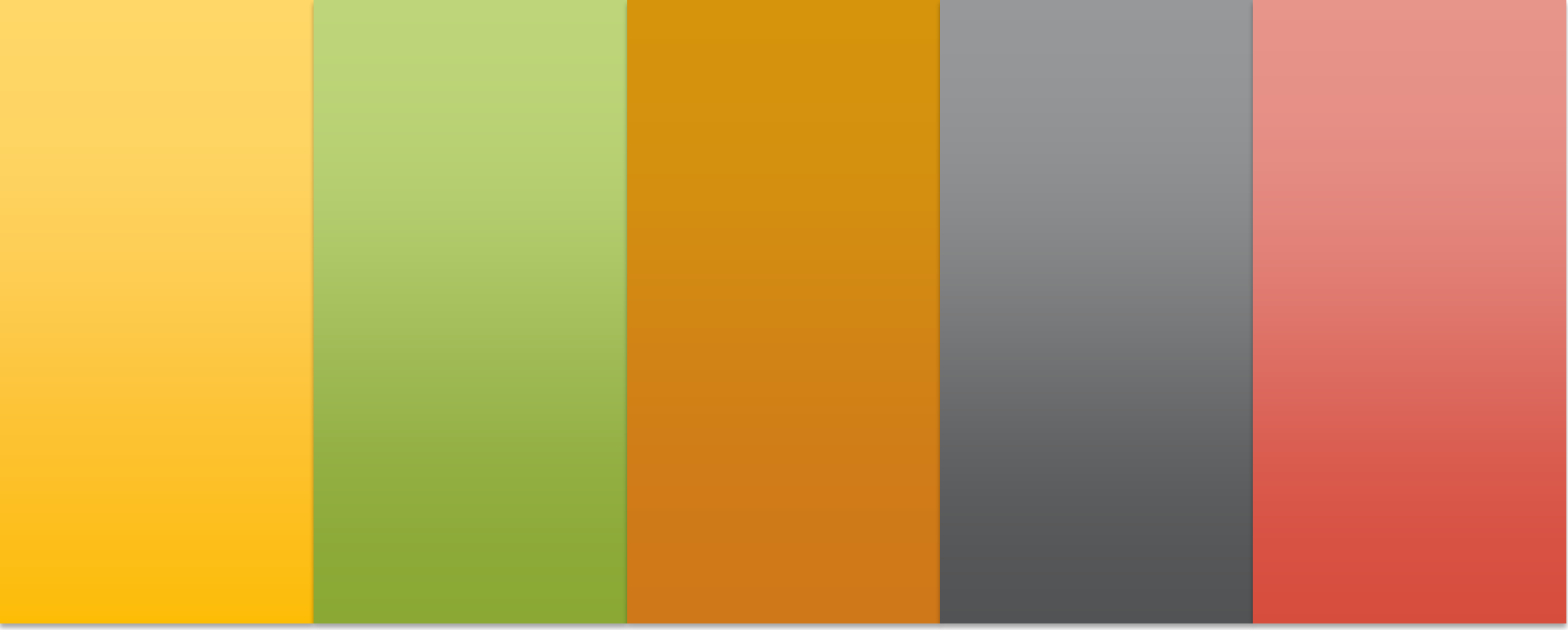
BIG DATA COMMITTEE - MISSION STATEMENT

- The challenges of traditional, sample-based market research continue to rise in our dynamic, fragmented media landscape. At the same time, accessibility to large data sets from almost any business or personal sector (including our own physiological responses) is driving increased demand and innovation within the data sciences.
- The goal of the Big Data Committee is to explore the growing intersection of these two disciplines, identify and create informed dialogue around the critical questions this intersection creates, and explore methods, techniques and approaches to improve the quality of big data solutions.

VENDOR PERSPECTIVE: Under The Hood Of Big Data

George Ivie, CEO, Executive Director, MRC
Ken Barbieri, VP Market Development, Neustar
Andrew Fiegenson, Managing Director, Nielsen
Mainak Mazumdar, Chief Science Officer, Simulmedia
Nishat Mehta, EVP, Global Partnerships, DunnHumby

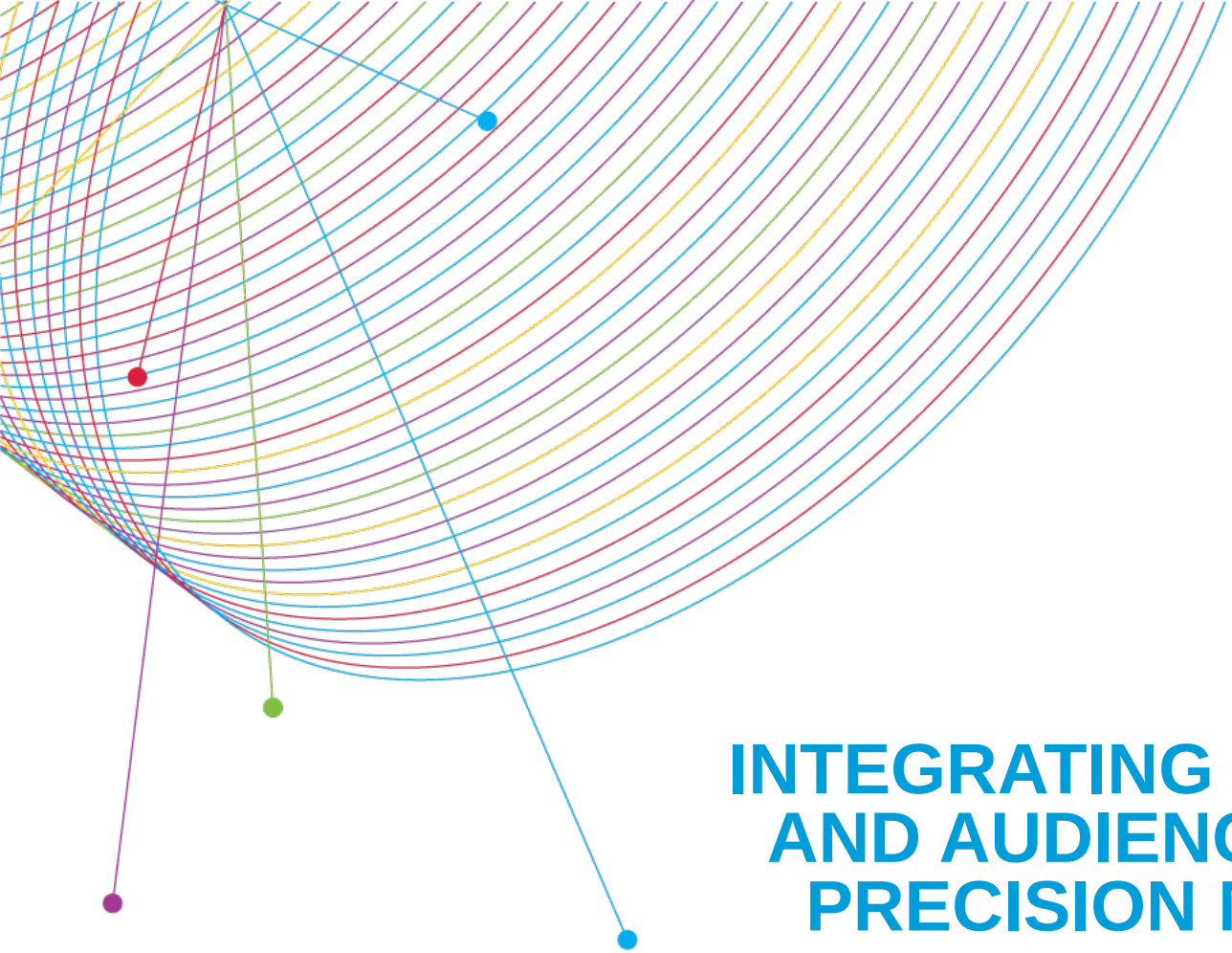
cre



BIG DATA CASE STUDY: Integrating Credit Card Transactions And Audience Data To Better Understand And Reach Consumers

Pete Doe, SVP Data Integration, Nielsen

cre



nielsen

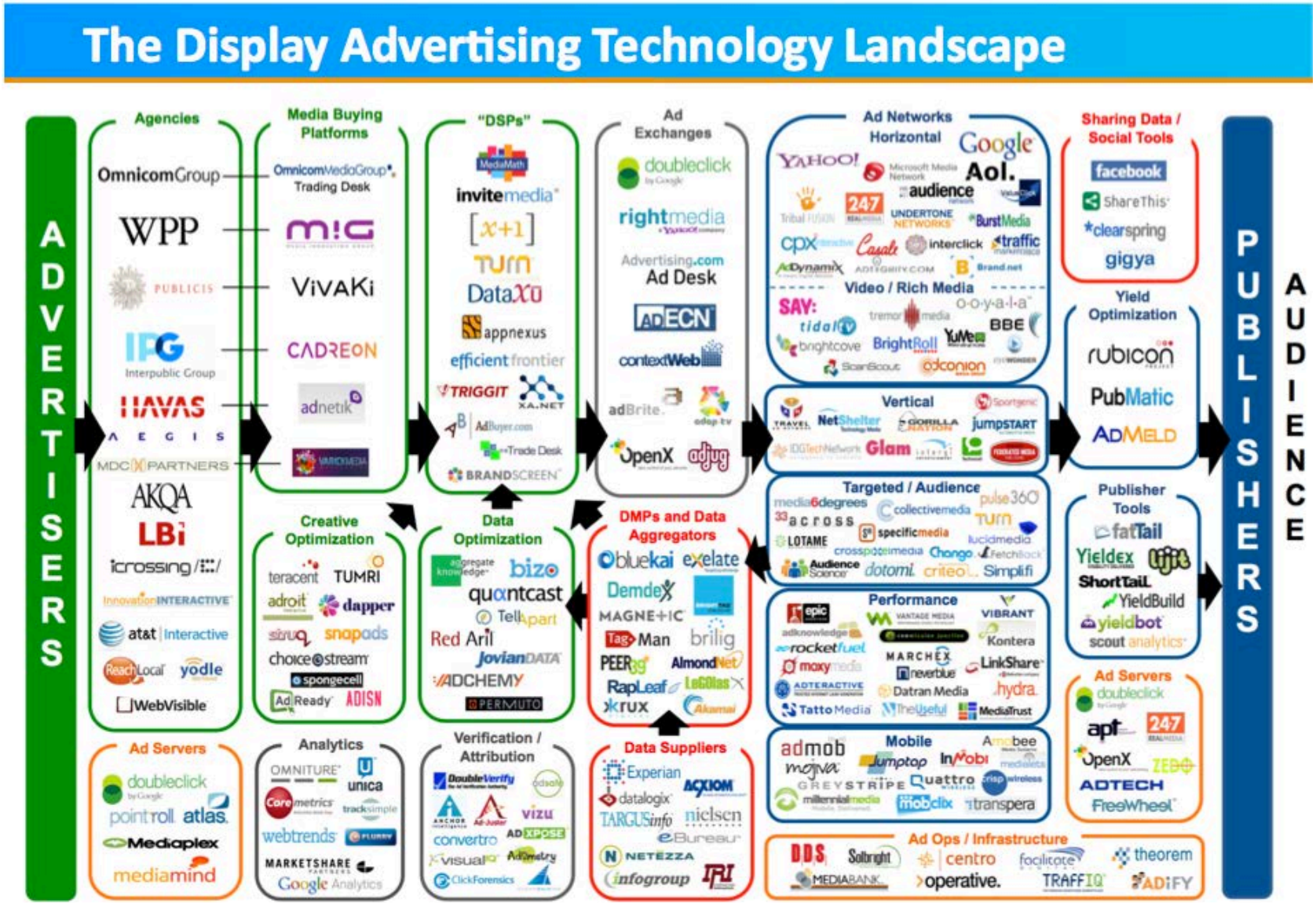
INTEGRATING CREDIT CARD AND AUDIENCE DATA FOR PRECISION MARKETING

PETE DOE

CONTENT

1. Background
2. The Opportunity
3. The challenges – Privacy, Integration Methods
4. Validity
5. Conclusion

BIG DATA HAS SPARKED AN EXPLOSION OF NEW TECHNOLOGY PROVIDERS



ONLINE AD LANDSCAPE

Online Advertising is increasingly moving to Programmatic Buying

Real-Time DAILY

Specific Media Expands Nielsen Partnership For Moviegoer, TV Audience Data

by Tyler Loechner, Yesterday, 2:35 PM

Specific Media on Wednesday announced it has expanded its partnership with Nielsen to include new sets of audience and segmentation data into its digital ad offerings, per a release.

Nielsen's "TV Audience Segments for Mobile" and "Movie Audience" segments -- which creates audiences based on moviegoing habits -- will both be made available to Specific Media clients.

Specific Media is the first beta media participant to use the TV audience segments for mobile, according to a release. The segments use data from Nielsen's National People Meter Panel and Online Panel behavioral data.

"Marketers are increasingly looking to build smarter cross-screen campaigns through precision marketing," stated Andrew Feigenson, managing director of digital client services at Nielsen.

Tags: ad technology, audience targeting, big data

OPPORTUNITY

Incorporating offline activity improves the efficiency of buying and selling

NIELSEN AND PROGRAMMATIC BUYING

Nielsen (and partner) data provides offline consumer and media activity

TV viewing

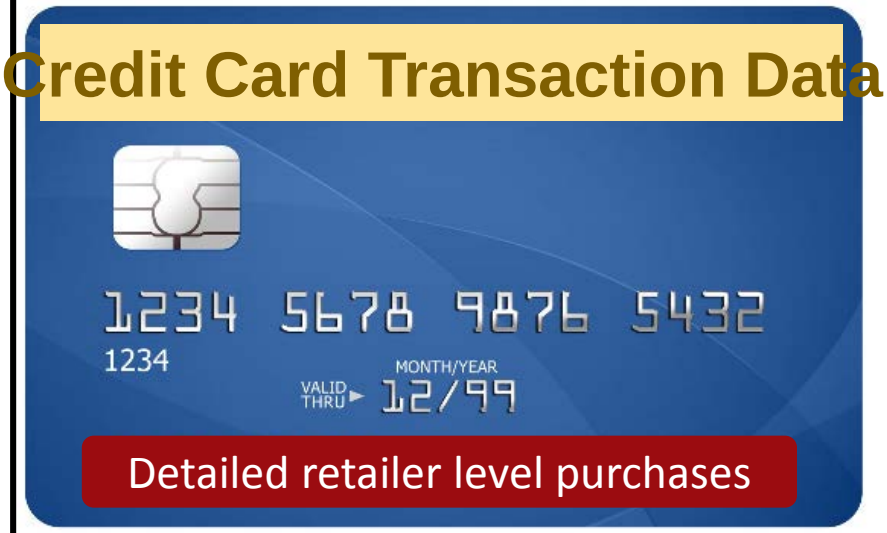
CPG segments

(Nielsen Catalina)

Prizm Segments

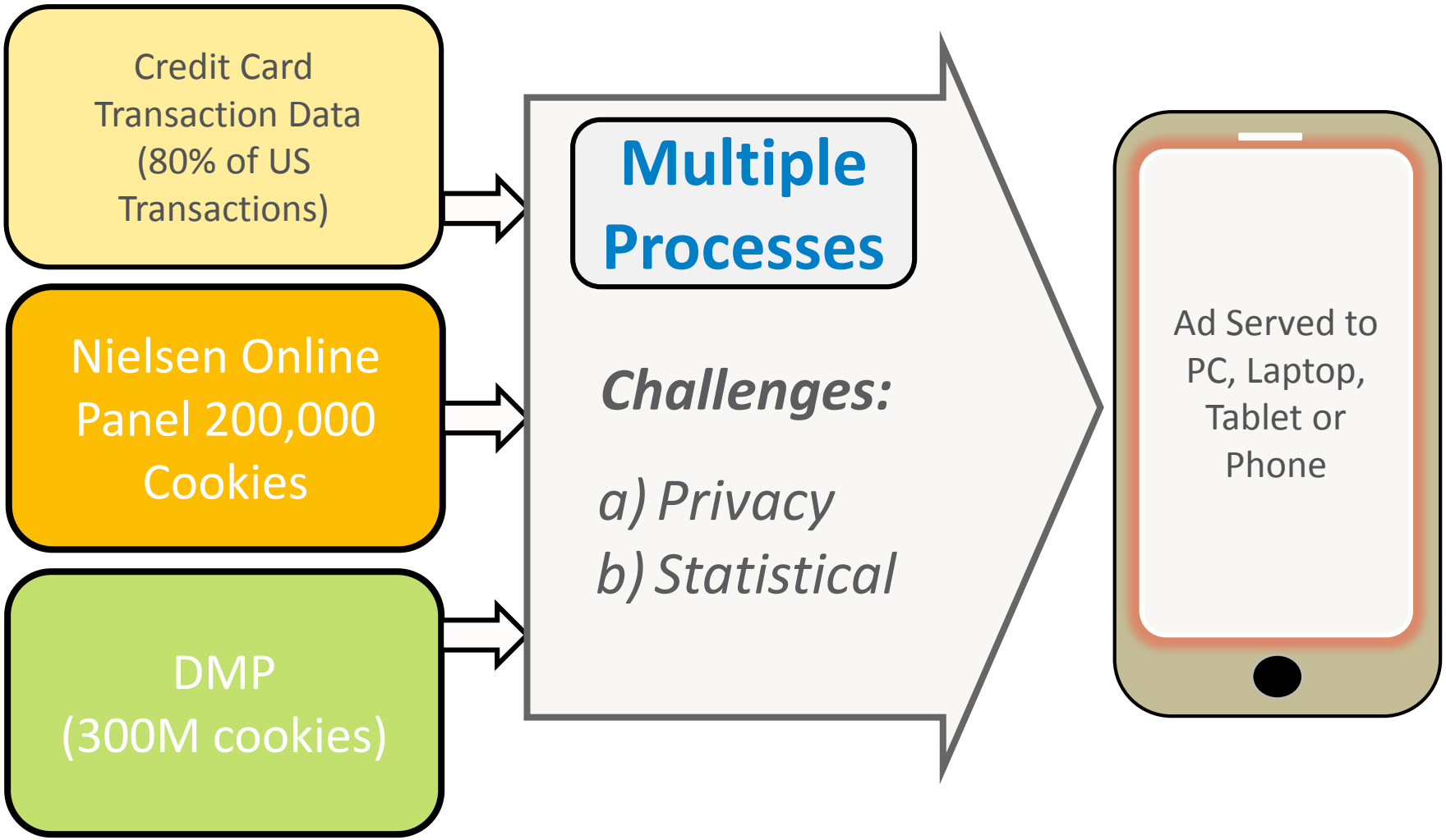
Credit Card Activity

Linking these data with online databases enables more effective online advertising

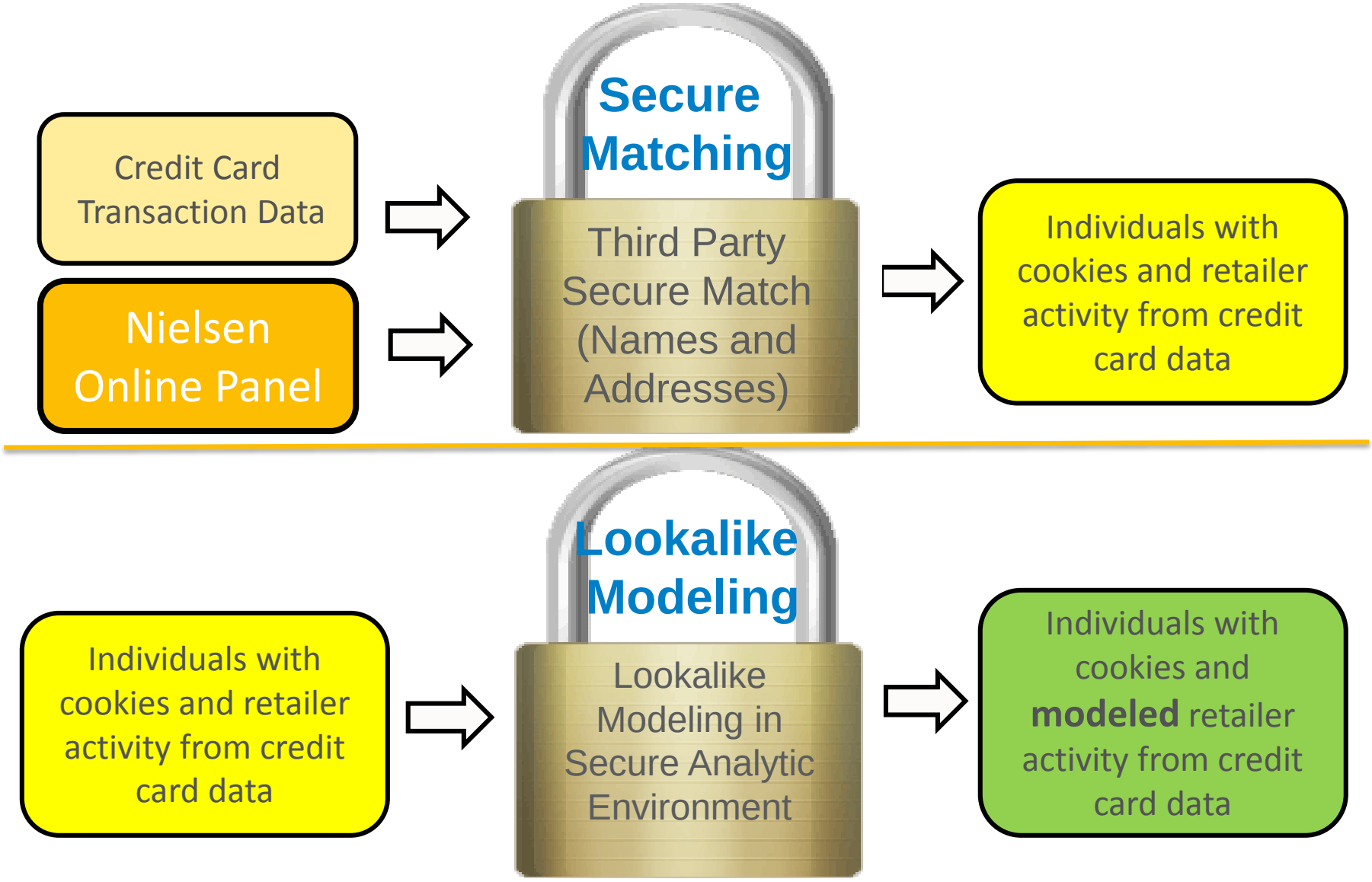


Airline	Hotels
Apparel	Mass Merchandiser
Baby Stores	Home Improvement
Book Stores	Pet Stores
Casual Dining	Supermarkets
Department Stores	Travel Services
Fast Food	Wireless
Fine Dining	Electronics

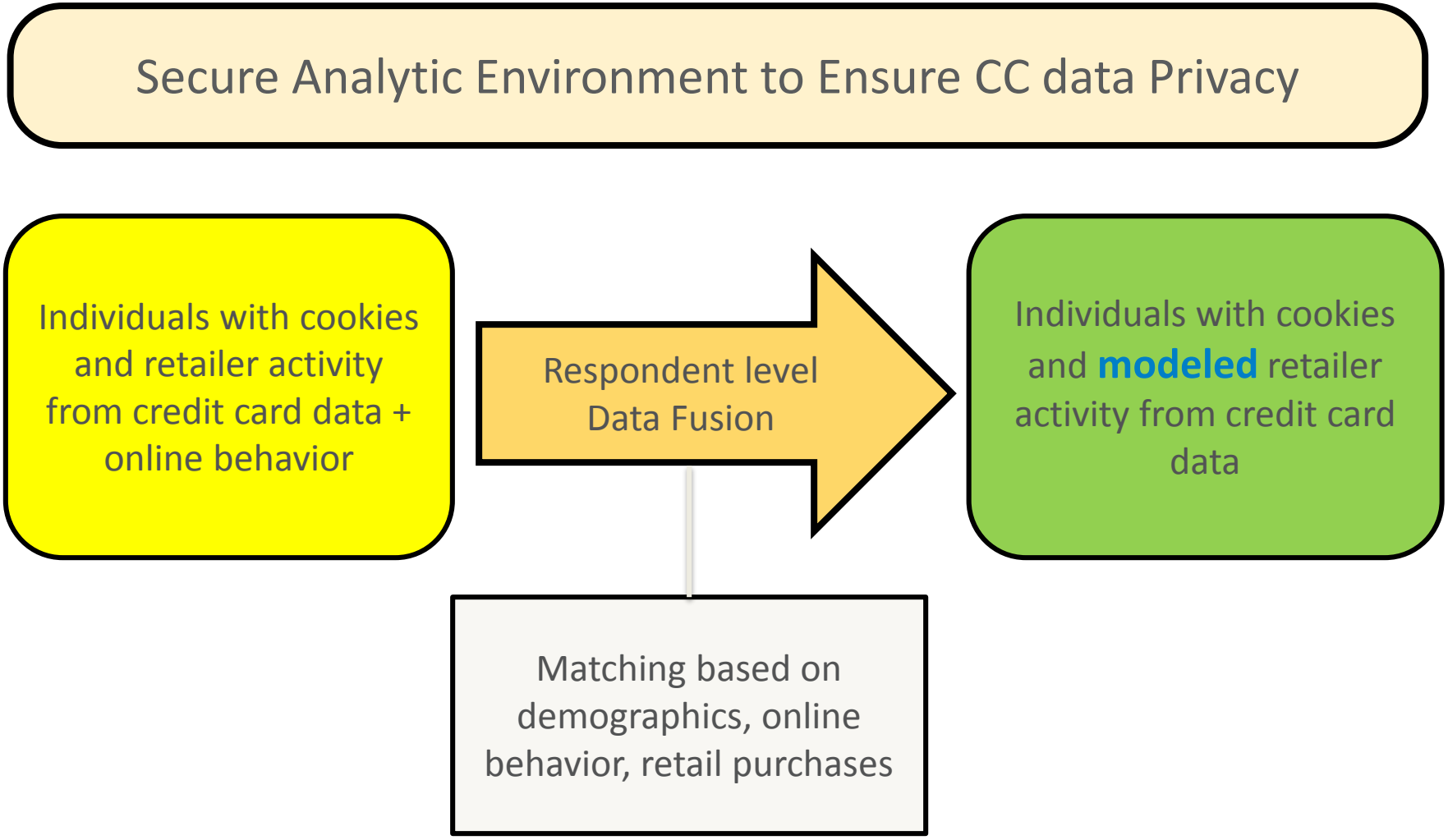
INTEGRATING RETAIL ACTIVITY INTO ONLINE AD SERVING



ENSURING PRIVACY – TWO PROCESSES



LOOKALIKE MODELING



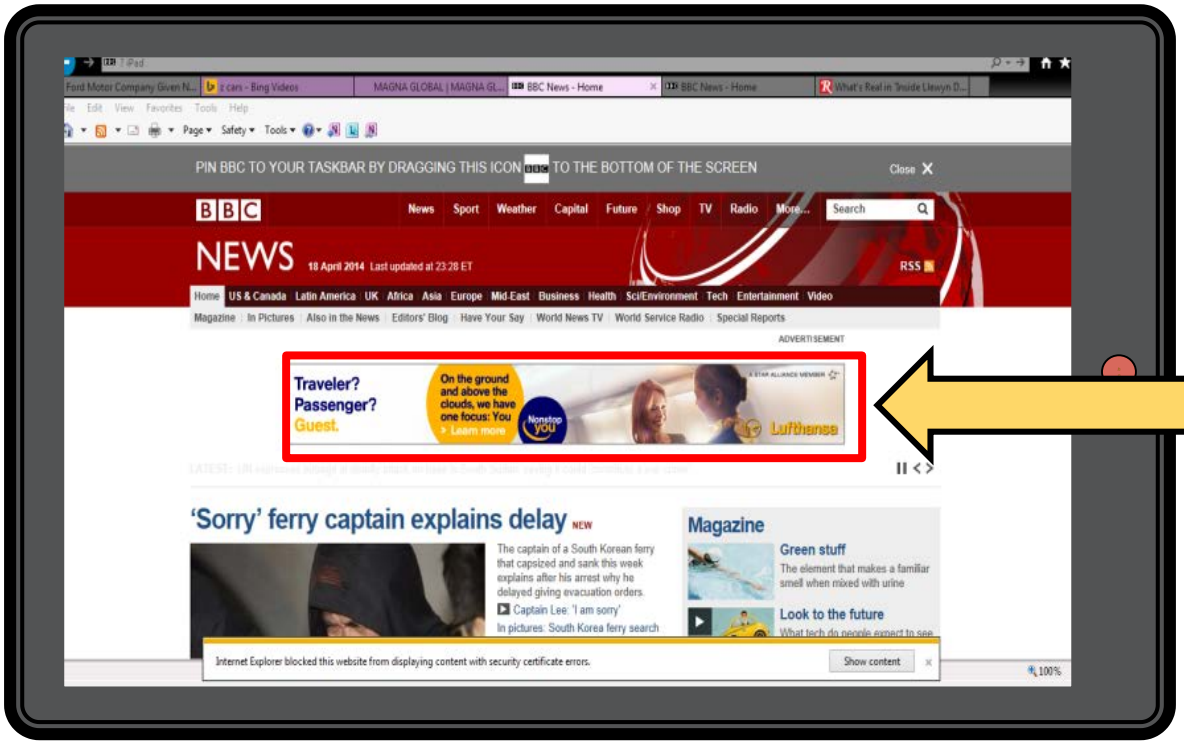
DELIVERING ADS

Individuals with cookies and **modeled** retailer activity from credit card data

Cookie Matching and modeling

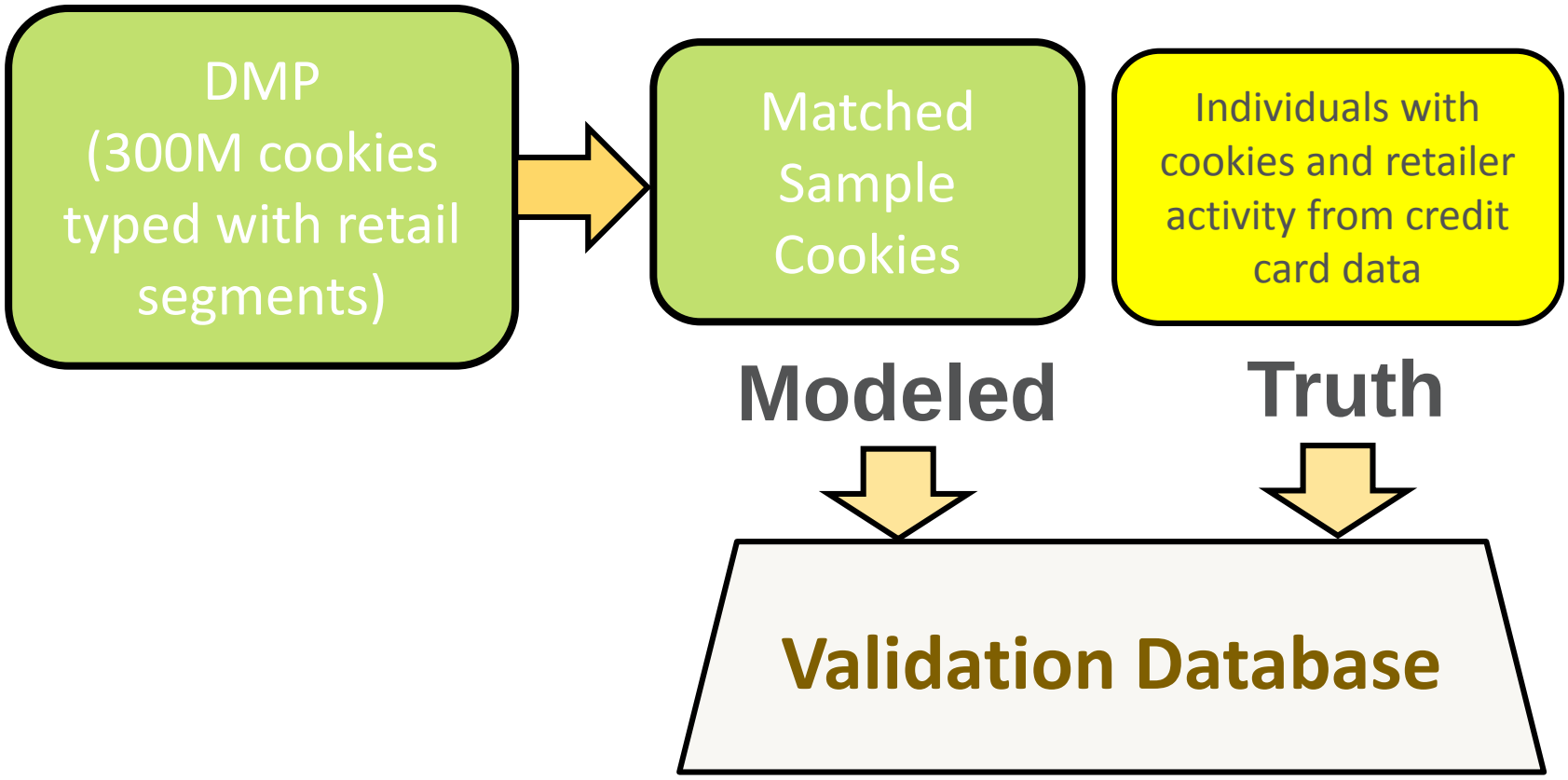
Modeling uses online activity tracked through cookies

DMP (300M cookies typed with retail segments)



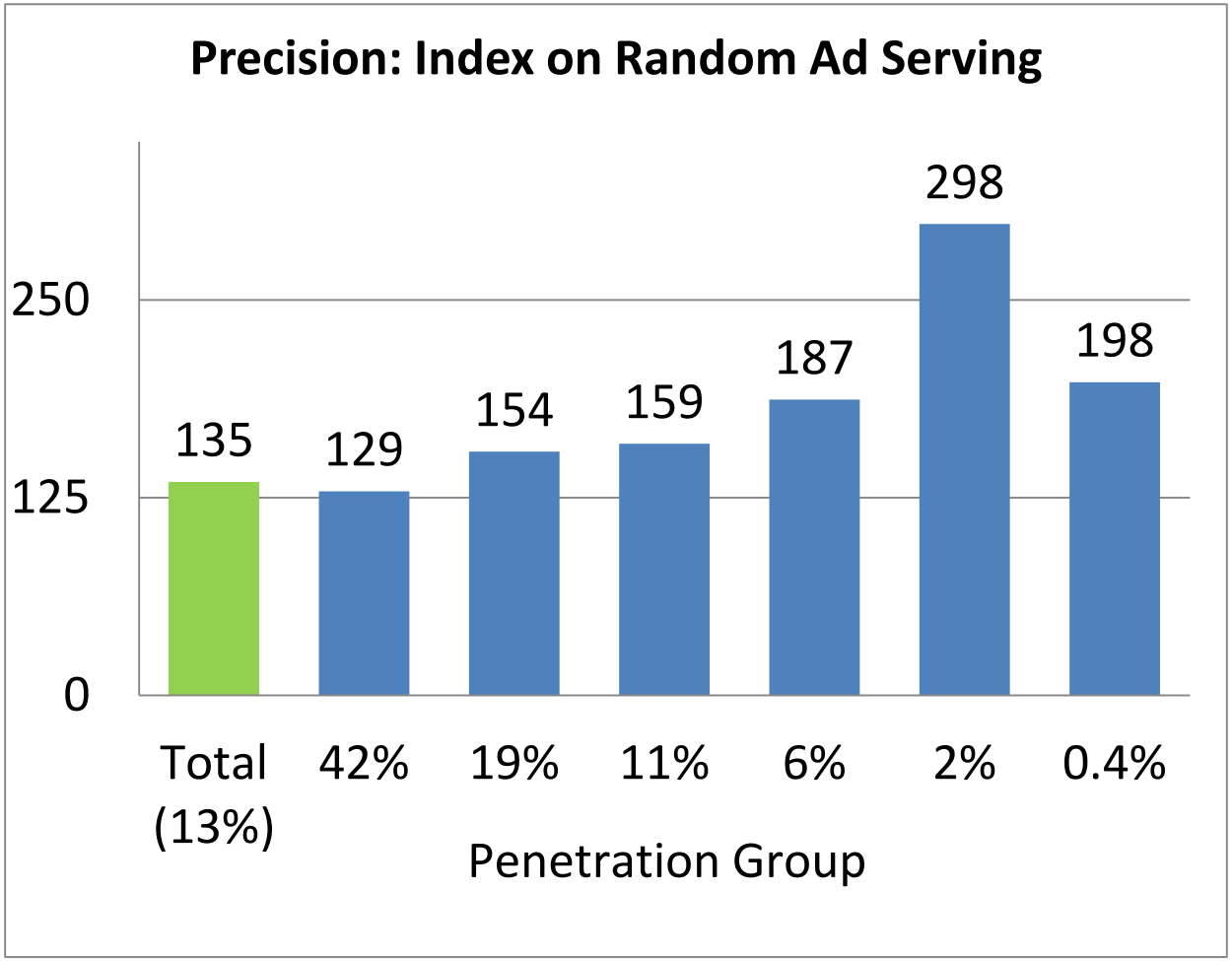
Ad Networks/ Publishers

STATISTICAL VALIDITY TEST



STATISTICAL VALIDITY

Our Objective: Demonstrate Precision via Improvement on Random Measure > 125

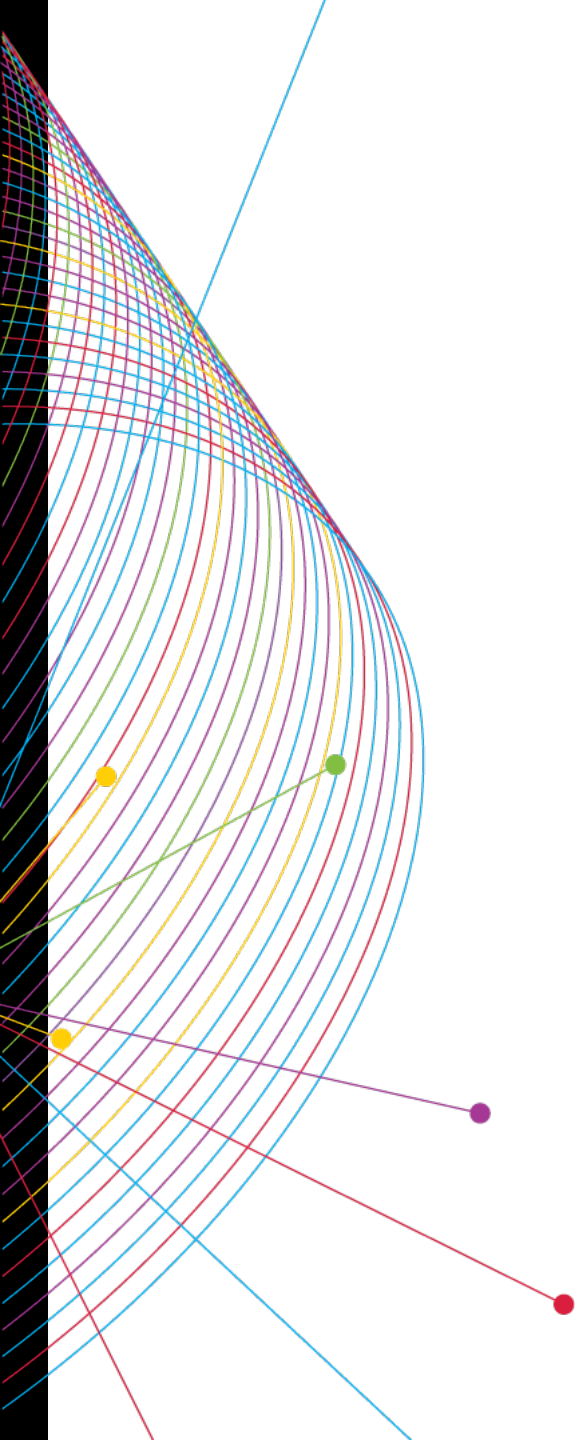


CONCLUSION

Making Ads relevant is good for the advertising industry and good for consumers

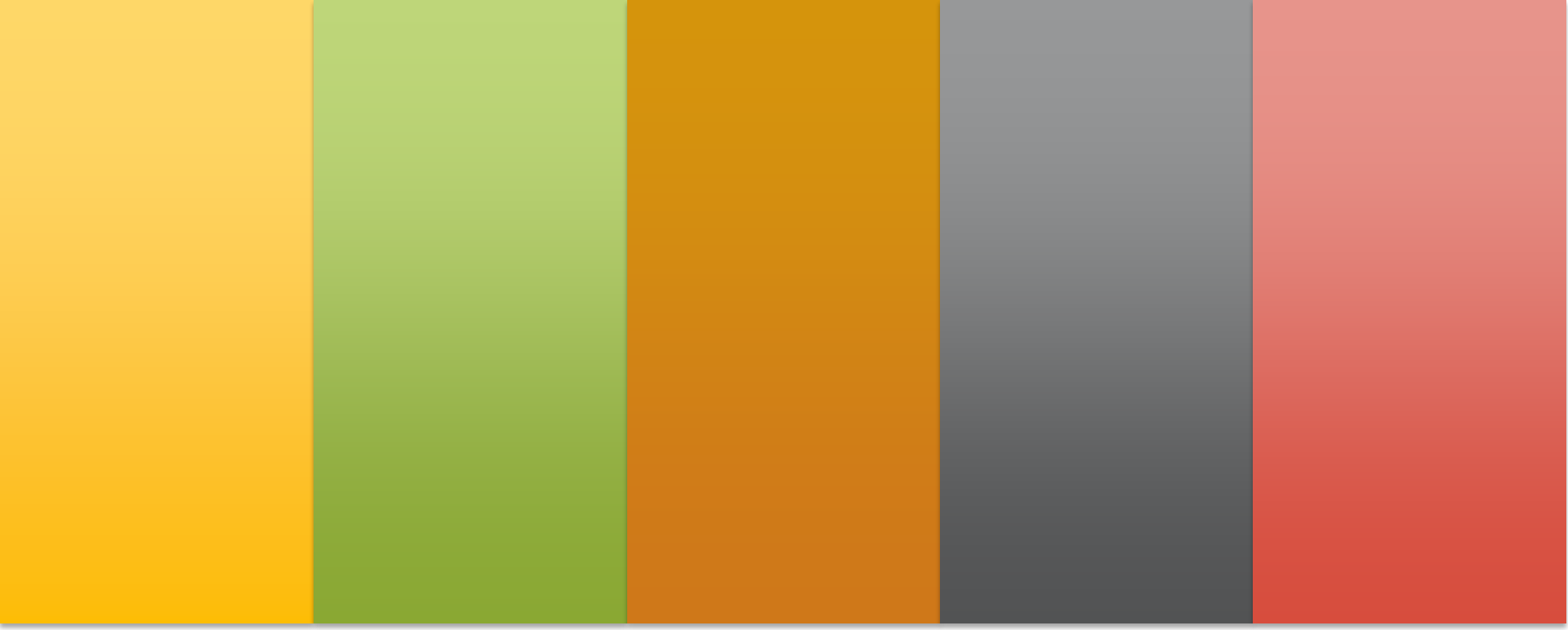
Integrating Data Sources delivers more effective and relevant online advertising

Good Data, Coverage, Privacy Compliance and Statistical Validity are Essential Elements



THANK YOU!

Pete.Doe@Nielsen.Com



BREAK

cre

MARKETER PERSPECTIVE: One Client's Data Journey

Mark Kaline, Former Global Media Director at Kimberly-Clark
and Former CRE Chair

cre

One Client's Data Journey

CRE Big Data Case Study Event – October 23, 2014

Mark Kaline

“By 2017, the CMO will spend more on IT than the CIO”.

Gartner, February 2012

> Big Data is commonly described by the technology industry using the 4 V's:

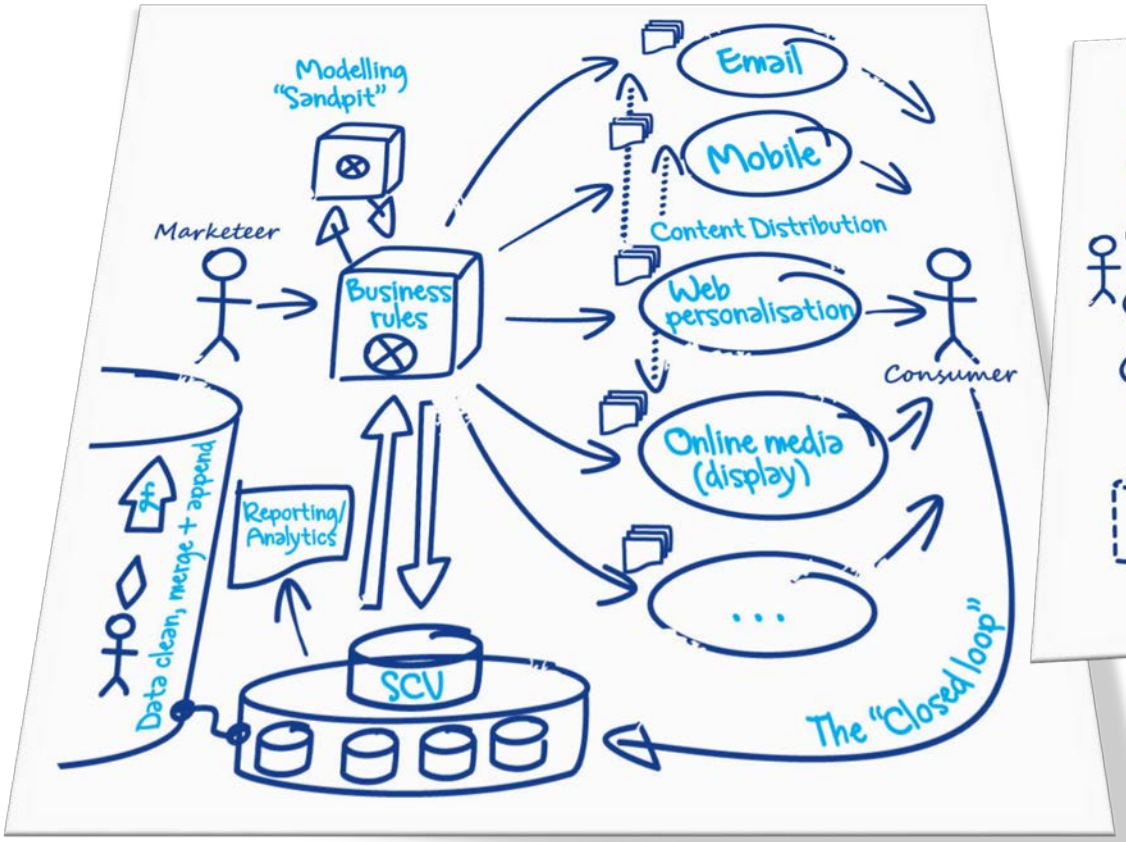
- Volume: Refers to the massive amount of data being collected by companies, through internal and external means
- Velocity: Refers to the frequency of data generation or frequency of data delivery (real time)
- Variety: Refers to the types of data being collected- structured (numbers, URLs,) or unstructured (video, images, text/chat)
- Value: Refers to the ability for the data to drive insights which could impact effectiveness, efficiency, profitability and growth.

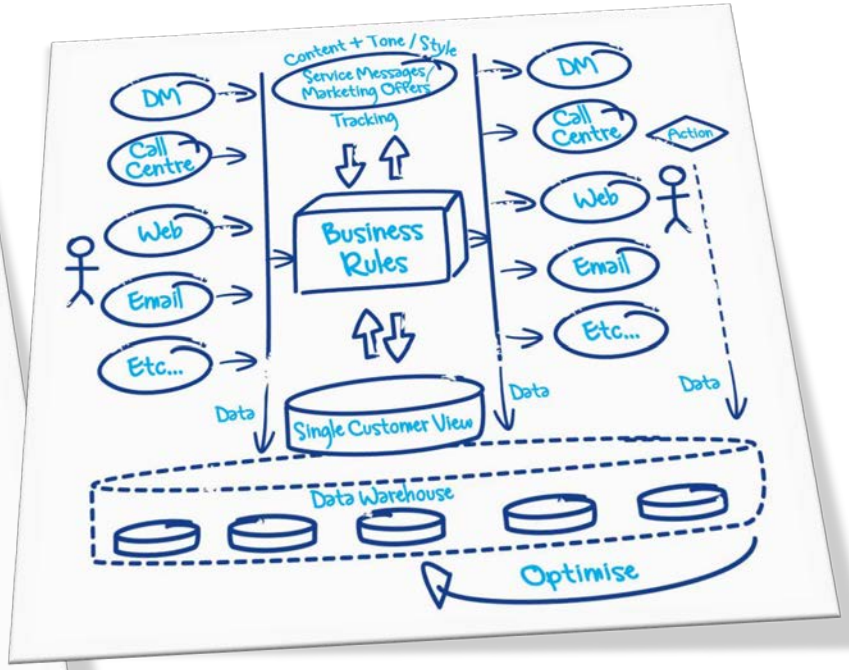
> Collecting Data is not enough:

- Most companies already have a flood of data – "*InfObesity*"
- Screen hopping and multi-tasking generating even more data
- What is it you are hoping to learn? What's the what?
- What are other parts of the organization hoping to learn?
- How does the organization collectively/exponentially learn?
- Do you have the right data? Is it good? And where is it?
- What is the plan to fill gaps in data to complete the desired analysis?
- Can it be repeated regularly & automated to speed learning?
- Does the data drive insights that are actionable?

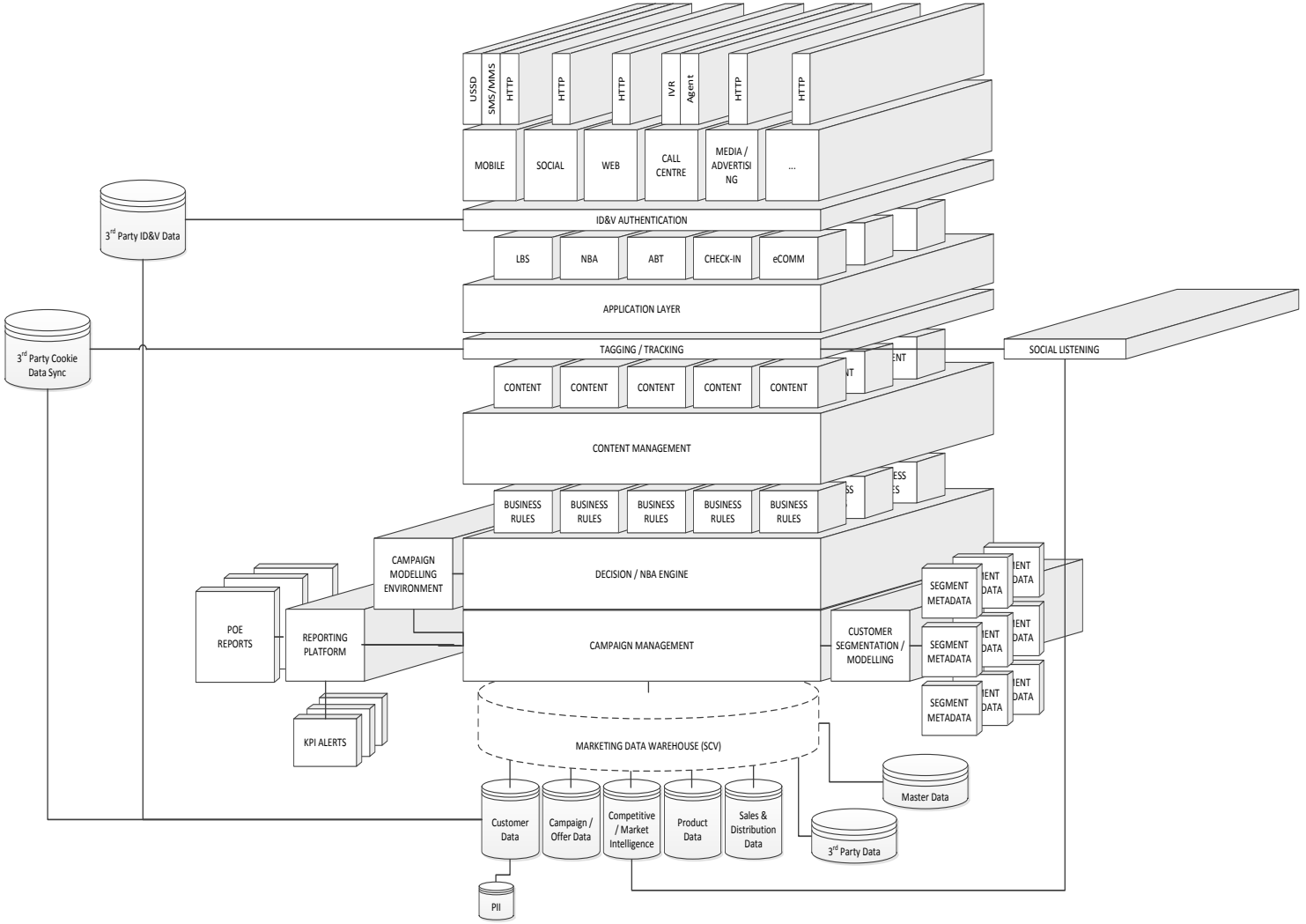
> Big Data is Nothing without Big Insights

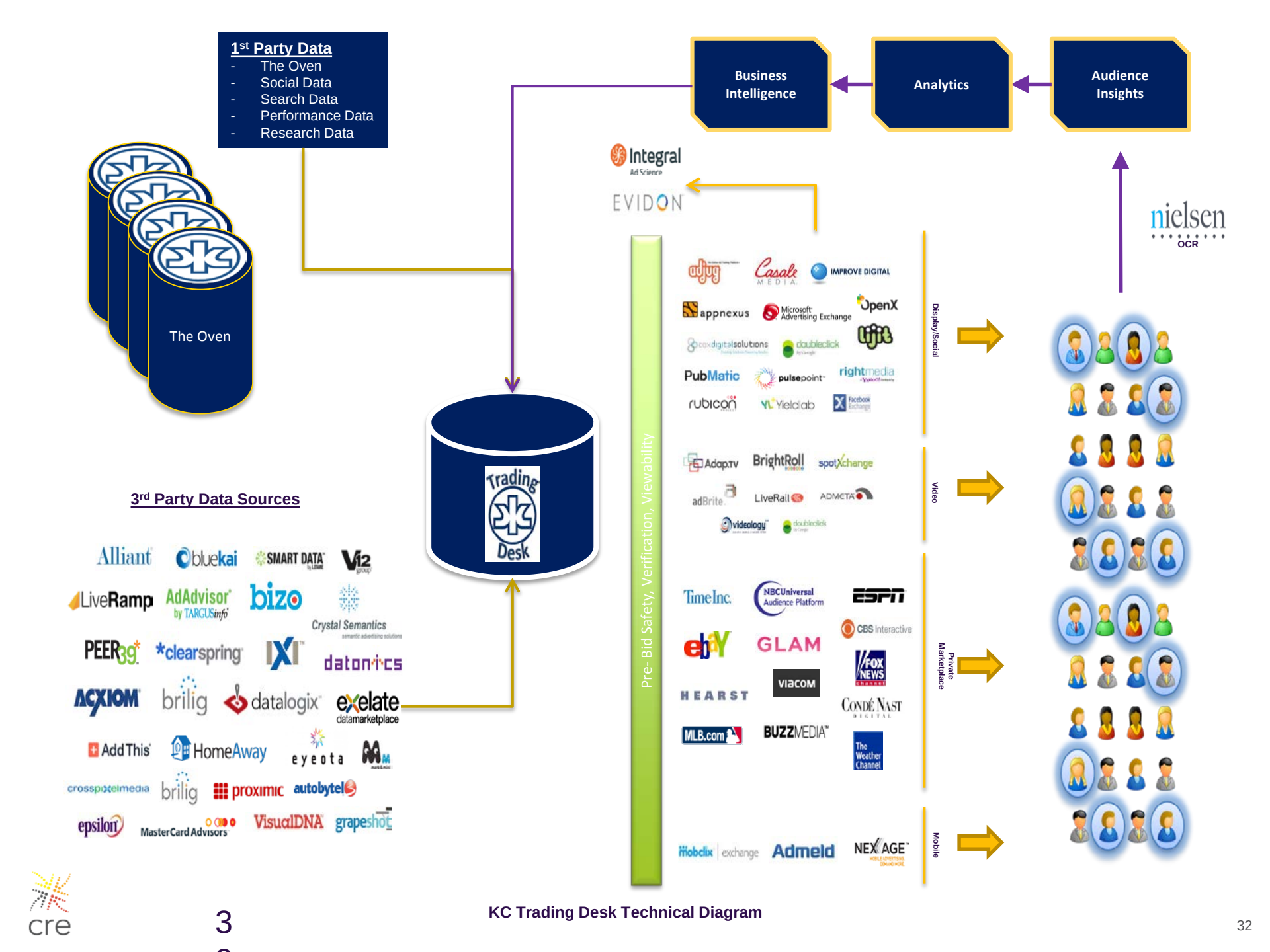
DEVELOPING AN APPROACH TO INTEGRATED MARKETING INTELLIGENCE





DEVELOPING THE TECH CAPABILITY FOR INTEGRATED MARKETING INTELLIGENCE





KC Trading Desk Technical Diagram

CURRENT KCTD DMP STATUS

- Over 5 million unique 1st party users
- Integrated with outgoing Oven Communications
- Connected to Social Media Activity
- Capturing Brand.com actions and personas
- Contains all Survey respondent information
- Linked to all Paid Search activity
- Actionable across all KC programs

KCTD DMP 2014 ROADMAP

Q1

Global Rollout

Deeper Integration with CRM

Test and Learn with Dynamic Creative

Q2

S2S Sync with OCR Video Provider

Q3

Connecting Offline & Online

Q4

Creation of Always-On Unified Targeting Profile

- Marketing organizations desire data & attribution to better drive decision making
- Clients can turn big data into a powerful media asset, driving optimization, increased ROI and targeting effectiveness.
- Moving from commercial programs that run, to commercial programs that learn
- Applying learning, where possible, in real time.
- Implementing programmatic buying where possible to learn dynamically.
- Recycling buy information thru CRM system to drive greater and greater precision.

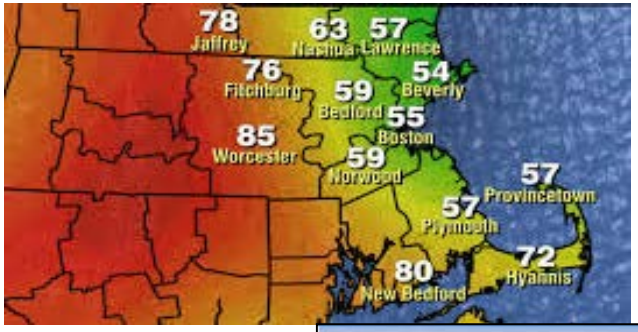
- Used criteria from CRM database, email, and websites to create unique targeting strategies
- Developed Strategic targeting and optimization architecture and rules
- Generated in highest conversion rate of any tactic within Potty Breaks program

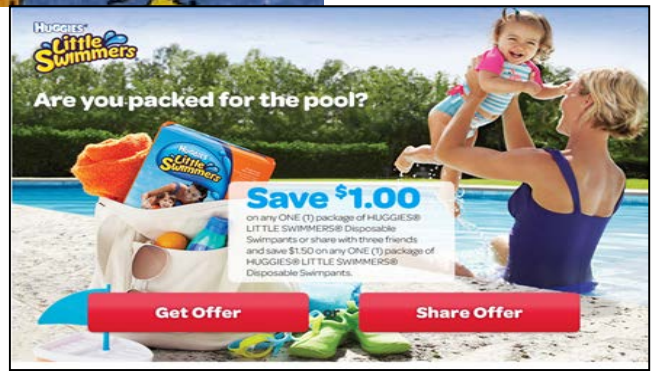
> **Challenge:** Swim diapers are not in season across the nation at the same time, so can waste be eliminated by focusing on markets that met certain criteria

> **Approach:**

- **Via Trading Desk,** use National Weather Service data feed to target geo locations that are 70+ degrees and sunny

> **Results:** Cost-per-action was **13%** below goal and survey results indicated **67% purchase intent**





- **Challenge:** Change the mindset from buying a single box of facial tissue during acute symptomology to buying multiples, **prior to the acute need**, in an effort to be prepared for the season.

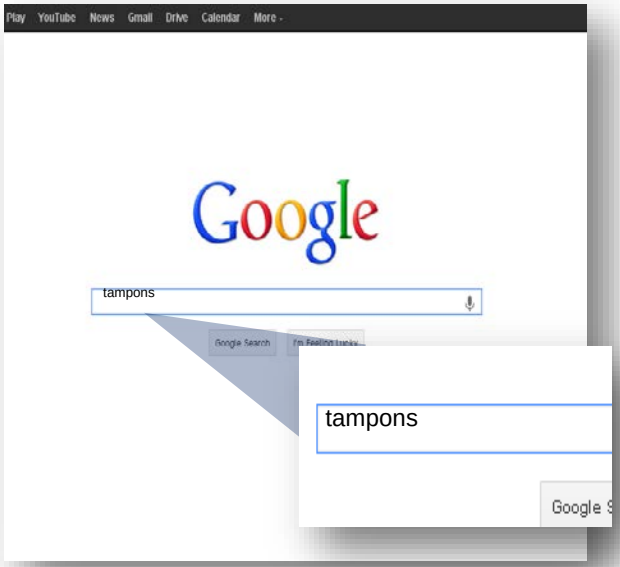
- **Approach:** Leverage Google's Cold/Flu data to geo-target our media in areas that have the highest Cold/Flu symptomologies

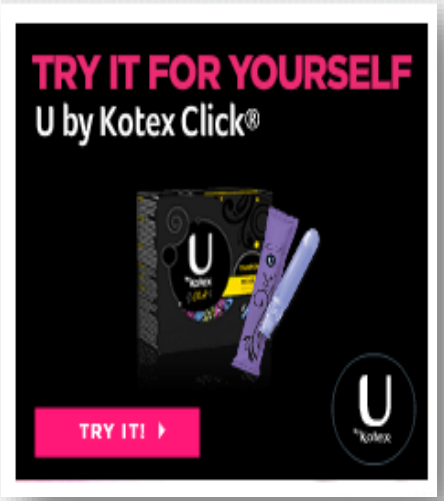
  - **Via Trading Desk:** Copy served based on level of flu symptoms in area
  - **Via Social:** Re-target those who mention cold/flu in posts

SEM RETARGETING

Retarget consumers searching specific keywords in Google
This tactic consistently performs among the best across various KC brands

SEM Retargeting ($7.52 CPA) outperformed all but 2 tactics in Kotex effort





User searches related keyword → KCTD captures user data → User targeted with UBK ad

DATALOGIX TV TARGETING

KCTD utilized data provider, Datalogix, to target consumers with a high propensity to have seen Snug & Dry TV ads

Process

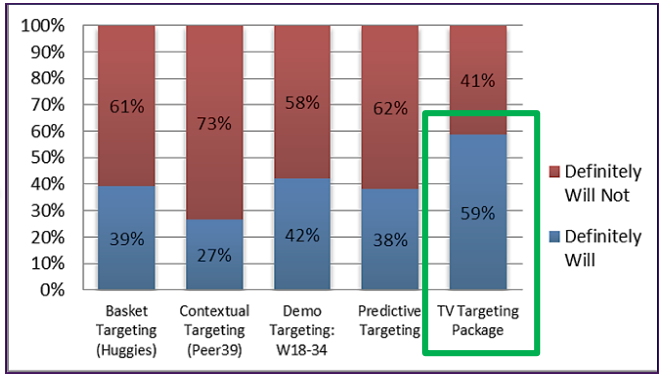
Consumer sees Snug & Dry Ad

Set Top Box Data collected; aligned with online addressable data

Consumer served Snug & Dry ad online

Result

TV Targeted audience was 59% more likely to purchase Diapers

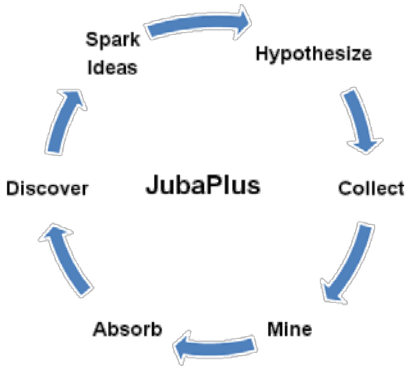


Q. How likely are you to buy Huggies Diapers in the next 30 days?

BIG DATA CASE STUDY: Can Behavioral Data & Machine Learning Algorithms Help Brands Grow Audience Interactions?

Yoram Greener, Founder, JubaPlus

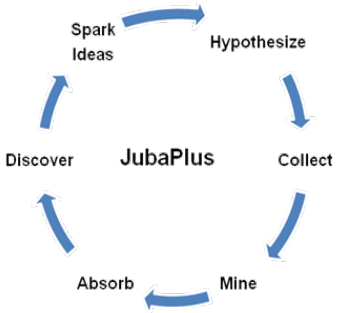
cre



cre
council for research excellence

http://bit.ly/P1OgH3

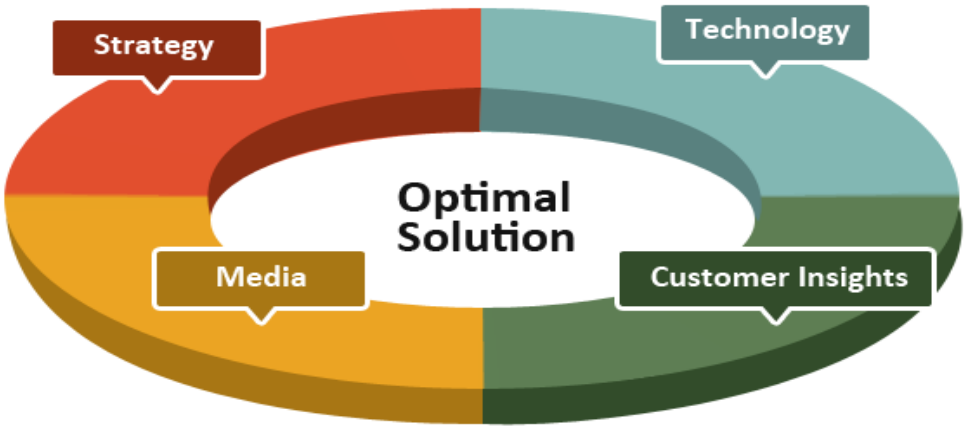
CAN BEHAVIORAL DATA AND MACHINE LEARNING ALGORITHMS HELP BRANDS GROW AUDIENCE INTERACTIONS?



AGENDA

- The Business Need: capture consumers moments of interests
- Industry Challenges
- Case Study

JUBAPLUS – AN OPTIMIZATION AGENCY



Strategy

- Identify customer growth opportunities
- Define the business case: Proforma P&L
- Mapping Marketing and Customer Goals to Business Goals
- Product Marketing Definitions

Media

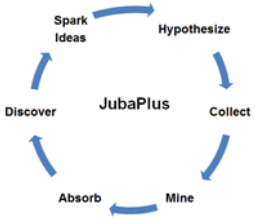
- Media Mix Models
- Media measurement
- Media Allocation & Budget
- Media Attribution
- Correlate Research Panels to Media Effectiveness

Customer Insight

- Digital Insights
- Social Listening Insights
- Facebook, Twitter and YouTube
- Ad-hoc Analysis
- Identify and outreach to Influencers
- Campaign Performance Reports

Technology

- Omniture and Google Analytics Tagging & Implementation
- Social Listening tools & NLP
- BI Tools & Dashboards
- Marketing Database Design, Hosting & Implementation
- Social CRM Technology
- Product Development



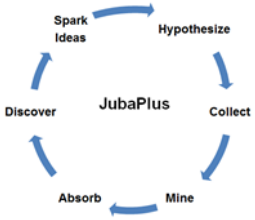
FULL CYCLE OPTIMIZATION FROM SPEND TO CONVERSION

Some 57% of marketers agree that data drives higher conversion rates, and 34% said it provides insights into customer behavior -- but most don't understand how to aggregate numbers from siloed media sources to drive overall better results

Failure To Master Online Data Costs Marketers Profits

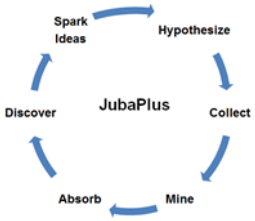
by Laurie Sullivan, MediaPost Jan 21, 2013

http://www.mediapost.com/publications/article/191528/failure-to-master-online-data-costs-marketers-prof.html?c=103844#reply#ixzz2JDwsBi4y



The Business Need
Capture consumers moments of interests

Consumers consume content across devices



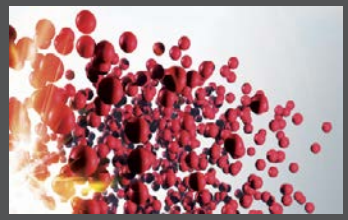
The Challenges: Information Access

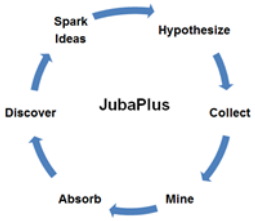
Consumers search and make decisions

- Faster than marketers can answer

- Anywhere 24/7

- Based on small but relevant particles of information





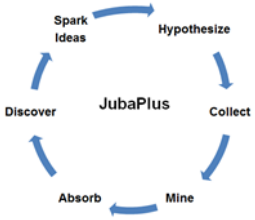
Industry Challenges: Classic Economic Problem of Demand and Supply

Too much consumer data, too fast

Consumer Demand for Relevant Information @ Speed

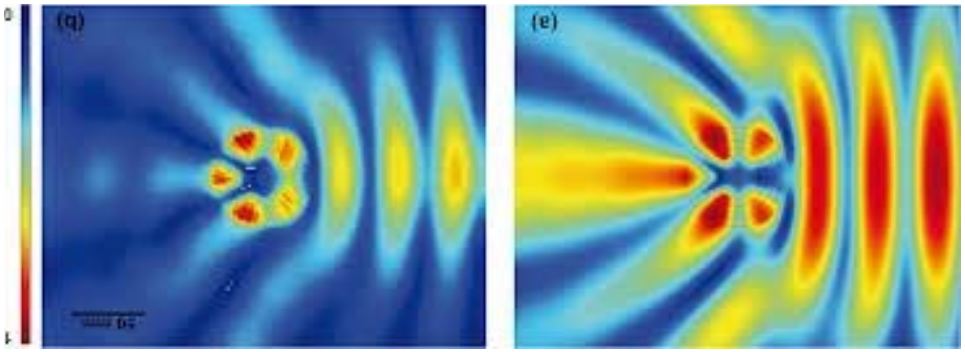
Brand Capacity and Cost

Marketers overloaded, over worked, while supply of content increases

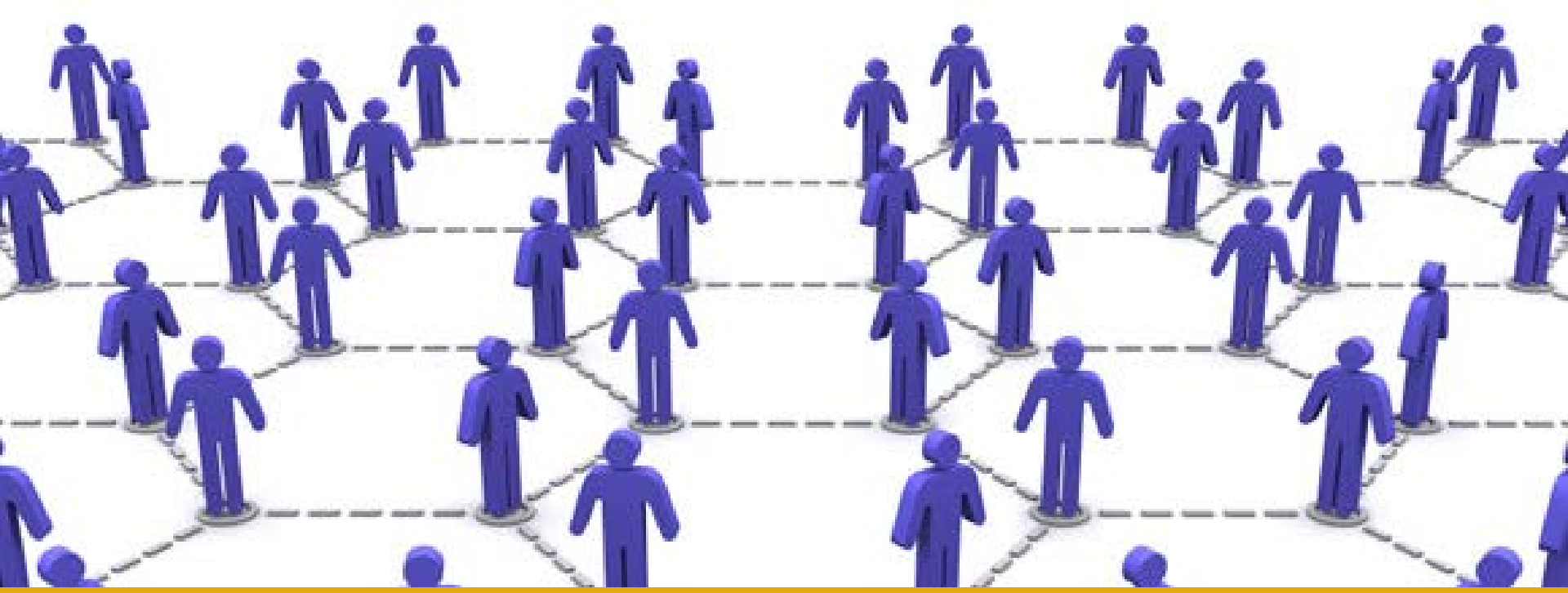


Industry Challenges: Classic Economic Problem of Demand and Supply

Finding the equilibrium point between consumer information demand and relevant content supply @ speed



Marketers have done sub-effective and deficient jobs: matching between what consumers search and their content servings



CASE STUDY: SOCIAL MEDIA CONTRIBUTION

November 2012

CONVERSATIONS AMONG CONSUMERS RESULTED IN XXX OF 3,088 ADDITIONAL XXX (3% OF 88,249 UNITS) AND 14,529 XXX (3% OF 415,130 UNITS) OR XXX OF $66.1 MILLION AND $435.8 MILLION OVER THE COURSE OF ONE YEAR.

IMPACT OF EARNED MEDIA – CONVERSATIONS AMONG CONSUMERS

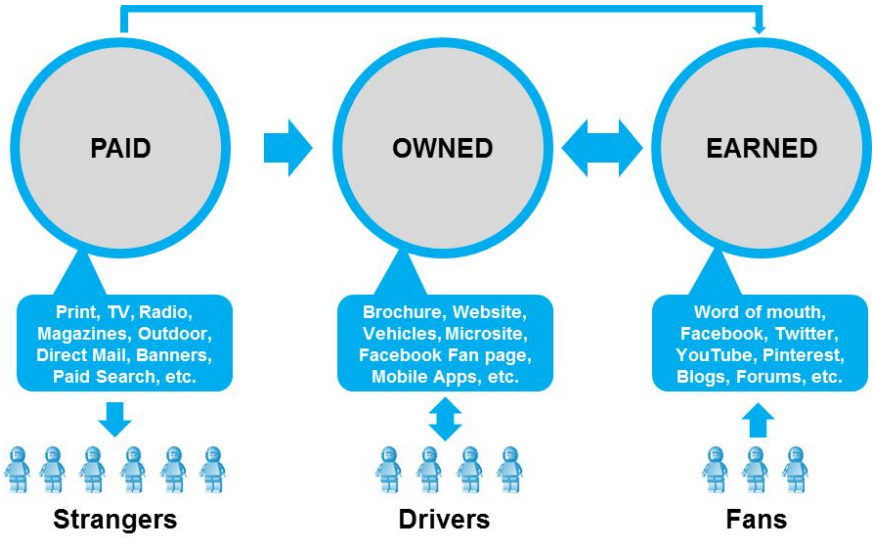
Objective:

- Quantify the impact of conversations among consumers on actual xxx

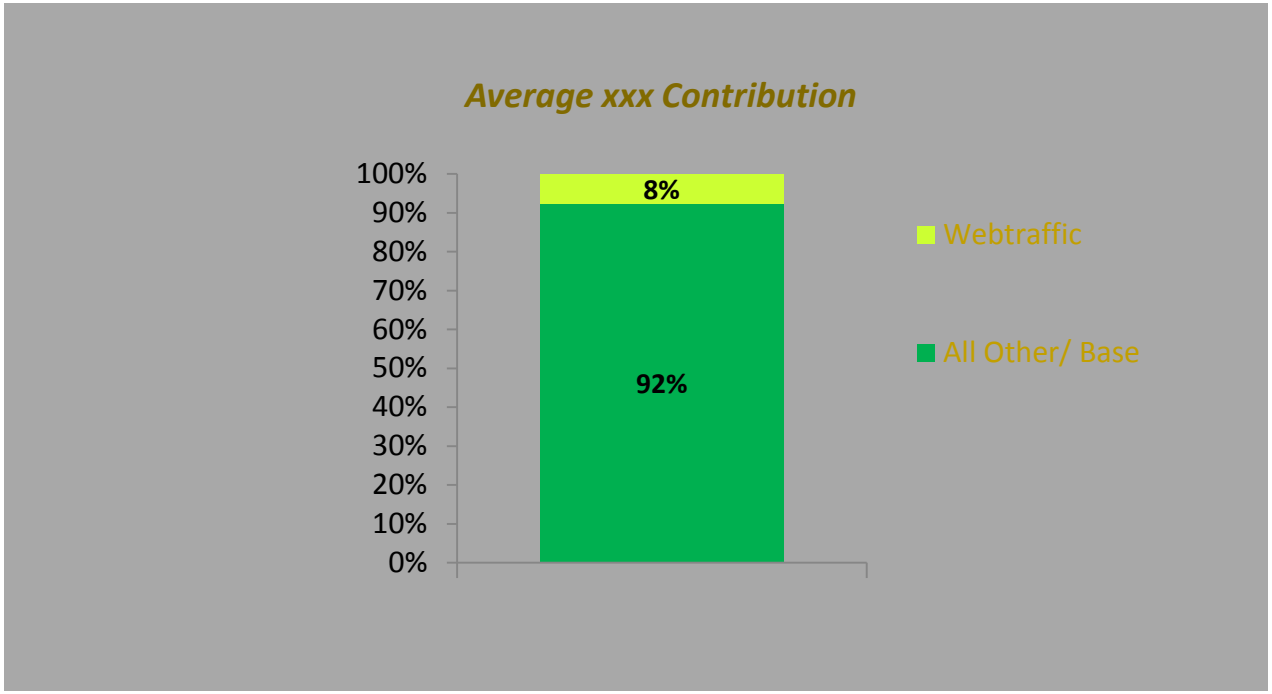
Methodology:

- Measure major media investments (paid, owned, and earned) between Oct 2009 and June 2011 on a weekly basis
- Use marketing mix modeling to isolate the impact of each media alternative (paid, owned, earned) on weekly sales
- Evaluate impact of alternative media types:
  - Digital **Paid** (online including search or SEM and display advertising on automotive sites such as Edmunds or Kelley Blue Book)
  - Traditional **Paid** (offline advertising, particularly nation-wide network TV)
  - **Owned** (branded Facebook pages, YouTube videos)
  - **Earned** (positive and neutral social conversations on social networks and blogs).
- Evaluate impact of sales channels (xxx events and xx activities, website visits)
- Earned Media and Owned Media are significant drivers of **Social Currency**

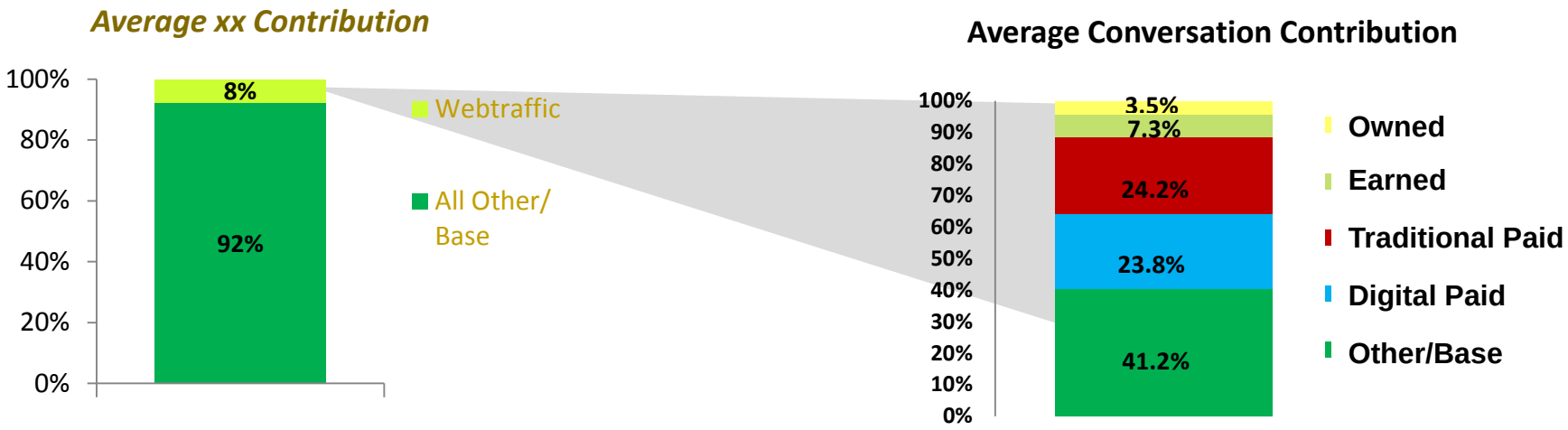
Alternative Media Types



OF TOTAL XXX OF 88,249 XXX AND 415,130 XXX, EIGHT PERCENT WAS ACCOUNTED FOR BY TRAFFIC ON THE WEB, WHILE THE REST WAS DUE TO OTHER EFFORTS SUCH AS XXX EVENTS AND XX ACTIVITIES

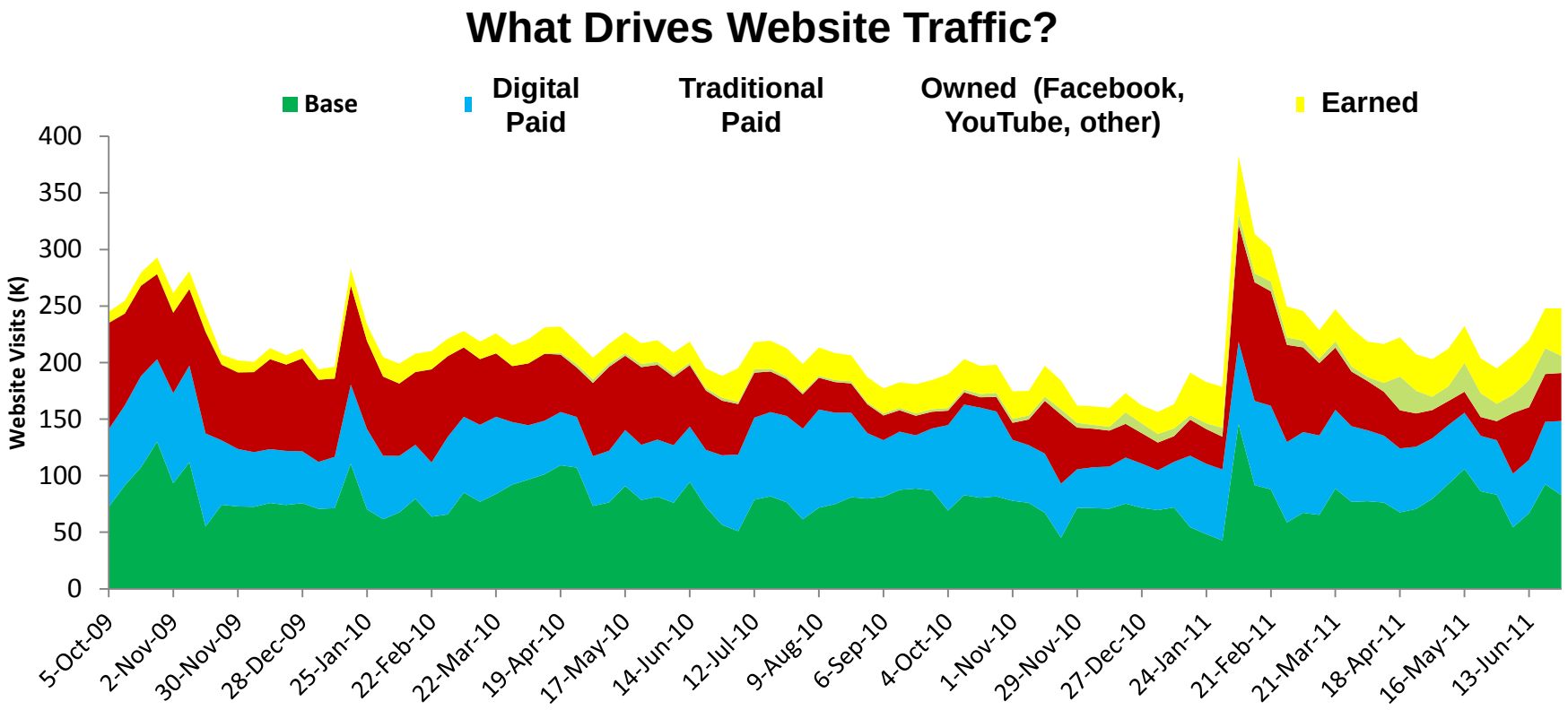


ON THE AVERAGE, TRADITIONAL PAID MEDIA CONTRIBUTED MOST TO WEB TRAFFIC THAT GENERATED EIGHT PERCENT OF TOTAL SALES; EARNED MEDIA CONTRIBUTED 7.3 PERCENT OF WEBSITE TRAFFIC



Base is the website visits that would be generated without Digital Paid, Traditional Paid, Owned, and Earned (Conversations) Media efforts

OVER THE COURSE OF 14 MONTHS, THE CONTRIBUTION OF DIGITAL PAID MEDIA WAS FAIRLY CONSTANT, WHILE TRADITIONAL PAID MEDIA (SUCH AS NETWORK TV ADVERTISING) DECREASED. OWNED MEDIA (VISITS TO FACEBOOK BRANDED PAGES OR YOUTUBE VIEWS) INCREASED GRADUALLY SINCE NOV 2010. CONVERSATIONS AMONG PEOPLE ALSO INCREASED GRADUALLY.



Base is the website visits that would be generated without Digital Paid, Traditional Paid, Owned, and Earned (Conversations) Media efforts

FINDINGS AND MAIN CONCLUSIONS

Overall Impact

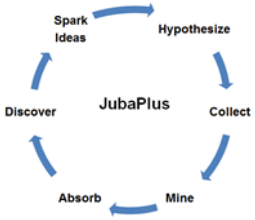
- Three percent of total xxx was driven by earned media (social conversations among consumers). With an average car model price of $25,000 for the xxx and $35,000 for the xxx, this means that social conversations contributed about half a billion in additional sales

Relative Importance of Media Types:

- Over the time period of this study, Owned Media and Earned Media had increasingly higher contributions to web traffic
- There was a significant increase in Owned Media (web traffic generated from Facebook / YouTube branded pages) that drove social conversations over digital paid and traditional paid network advertising

Our main conclusion is that the profits realized from $66.1 million and $435.5 million of additional xxx generated from social media channels (owned media on Facebook or YouTube, etc.) or through social conversations significantly outweigh the investments.

As far as marketing communications was concerned, the investment to profit ratio of social media was among the most efficient of all media alternatives



THANKS

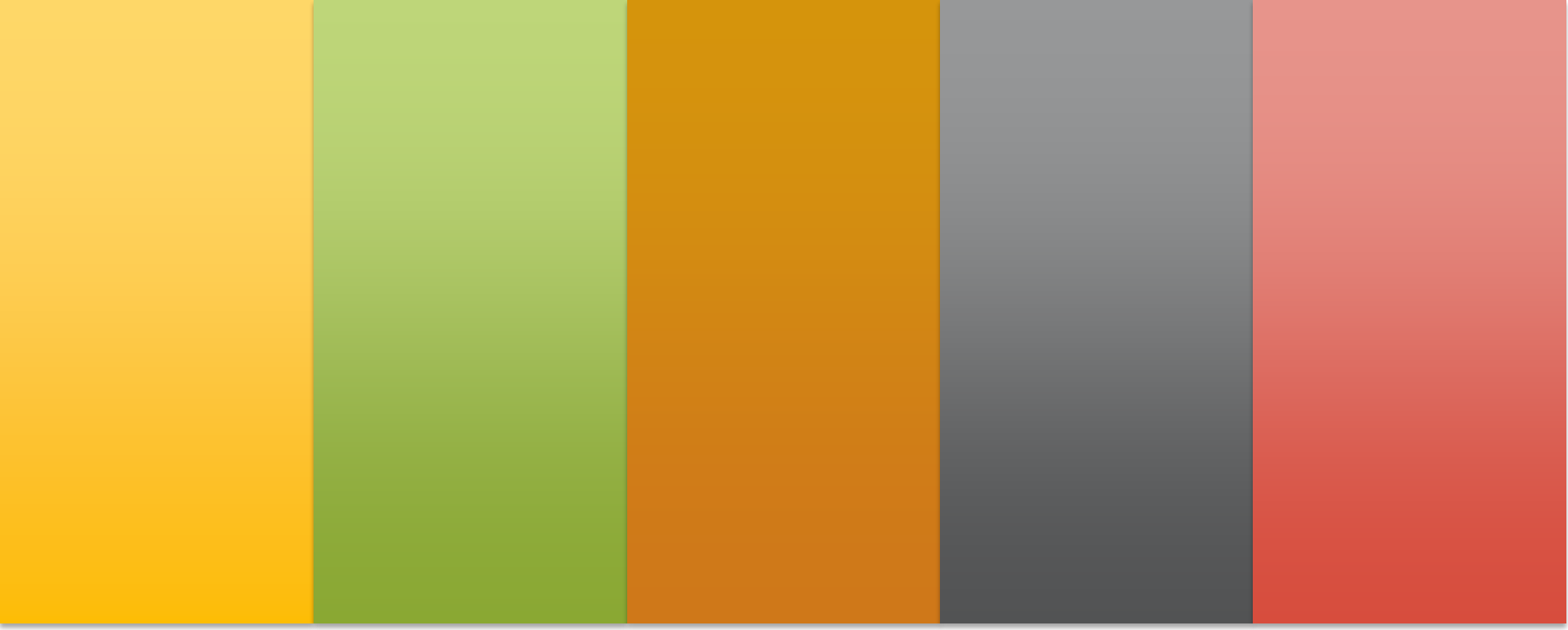
Yoram Greener
yg@jubaplus.com

Jubaplus.com
@jubaplus

MEDIA PERSPECTIVE: Putting Big Data In The Media Kit

Shaun Doyle, CEO, Cognitive Box
Howard Shimmel, Chief Research Officer, Turner Broadcasting

cre



CRE JUMPS INTO BIG DATA

Richard Zackon, Facilitator, CRE

cre

CRE PROJECT REPORT
A PREDICTIVE MODEL OF LOCAL TV RATINGS USING SUPERVISED MACHINE LEARNING.

CRE Big Data Event

October 23, 2014

PROBLEM

> Over 150 local TV markets are currently measured by Nielsen using only a paper diary.

> CRE has demonstrated that relatively small sample sizes render audience ratings unstable and nonresponse bias further compromises the accuracy of diary estimates.

> Broadcasters and advertisers in small markets lack reliable and valid metrics with which to plan and conduct business.

OPPORTUNITY

- The Nielsen National People Meter has some 20,000 HHs and is adding an additional 2300 households in diary-only markets.
- By applying techniques of machine learning, it may be possible to use these data to estimate local market ratings with accuracy significantly greater than current diary sample methods.

CRE LOCAL MEASUREMENT COMMITTEE

- Under the leadership of Billy McDowell (Raycom Media) the committee is exploring alternatives to paper diary measurement.
- They have commissioned a team to explore whether machine learning techniques can improve the accuracy of ratings estimates.

RESEARCH TEAM:

- Vasant Dhar: NYU Stern School of Business. Data Scientist
- Tim Dolson: LORE Media Research (Formerly VP, Statistical Methods, Nielsen). Data Consultant
- Sandy Retsky: Independent. Database Programmer
- Richard Zackon: Audience Patterns LLC, Project Manager

SUPERVISED MACHINE LEARNING

- An artificial intelligence technique in which the computer is presented with example inputs and their desired outputs in order to learn a general rule, an algorithm that maps inputs to outputs
- Inputs: TV Household characteristics, TV station characteristics and demographic viewing in 20,000 People Meter Households from outside a local market.
- Outputs: ratings estimates for the specific local market.
- Machine learning will develop a fitted model based on a “training set” and assess predictive accuracy with a “test set” of 60 stations.

RESEARCH PLAN

- A test of 60 stations in ten Local People Meter markets
- Demos: HH, P2-17, M18-49, M50+,W18-49, W50+
- To predict local ratings by QH ratings for 16 weeks in 2013 using simultaneous People Meter data from outside each market
- Compare predictions with actual People Meter estimates from within the market to assess the accuracy of the predictions

DATA

- Nielsen has provided tuning and viewing data by QH for Households and persons for all People meter HH's for February/May/July/Nov 2013.
- Nielsen has also provided Universe Estimates by market and demographic, geographic and psychographic data (e.g. Claritas) for each sample household.

ANALYSIS

- The research team will prepare the data for analysis by assessing its quality, analyzing distributions and performing appropriate transformations.
- This will be followed by assessing various machine learning methods for the problem and applying the resultant algorithm on data from respondents with known features.
- The test metric will be Live Demographic ratings by QH.

DELIVERABLE

> The CRE will be provided with a final report which will describe the analytic process including algorithms, validation statistics and recommendations for further R&D improvements to be taken up by CRE or Nielsen.

TIMETABLE

> September 4: CRE Approval
> October 6: Final specs
> October 10: Nielsen provided initial data
> November 14: Initial progress report to CRE Local Committee
> December 15: Final report to full CRE

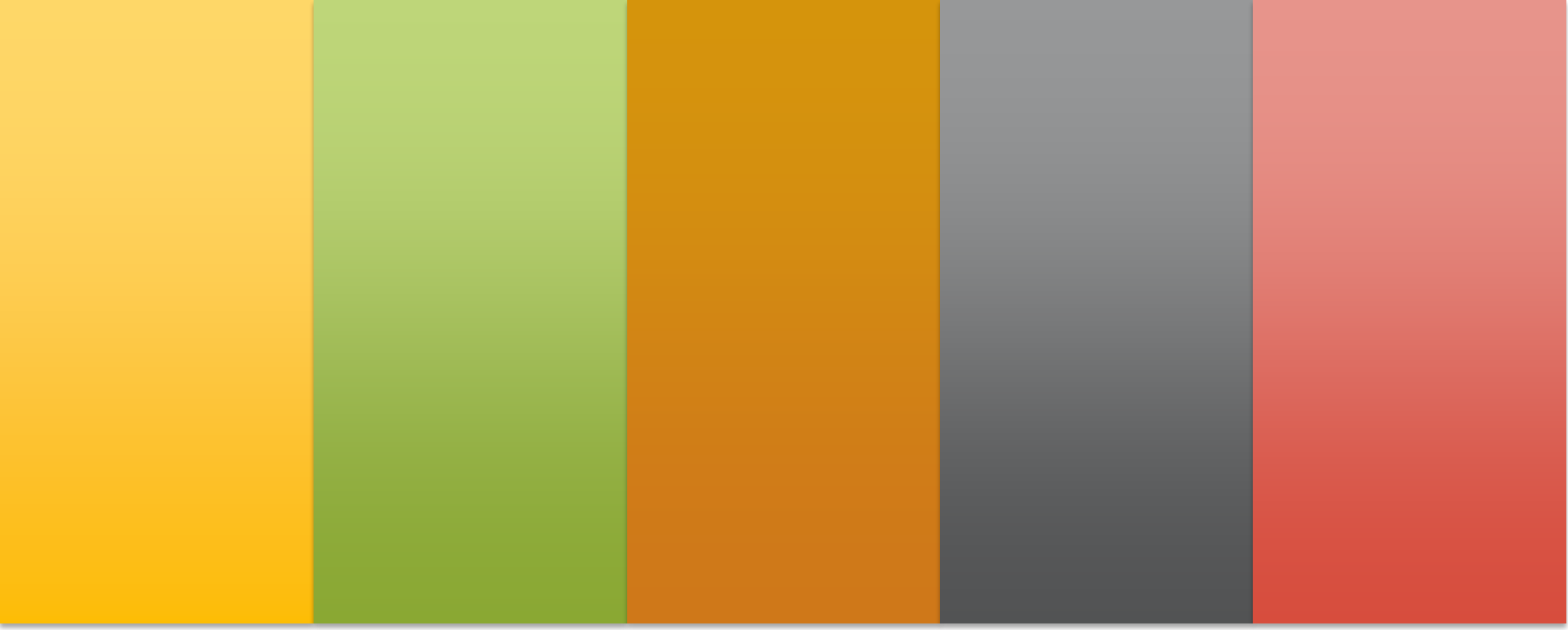
BENEFITS

- **If the results are successful:**
  - Local TV markets can consider an affordable new currency, more stable than one based on small samples and more valid than one based on diaries or STB's.
  - Nielsen will be encouraged to consider further refinements to the model we develop.
  - The industry will be encouraged to innovate with careful use of advanced analytics.
  - The industry will have had a front row seat to learn from the project.

- **If the results are unsuccessful or are inconclusive:**
  - The industry will have established a benchmark level of predictive accuracy for machine learning techniques, setting the basis for further improvement.
  - The industry will have had a front row seat to learn from the project.

CURRENT QUESTION:

What is the standard of accuracy required?



INTRODUCING: CRE Big Data Primer

Gerard Broussard, Principal, Pre-Meditated Media

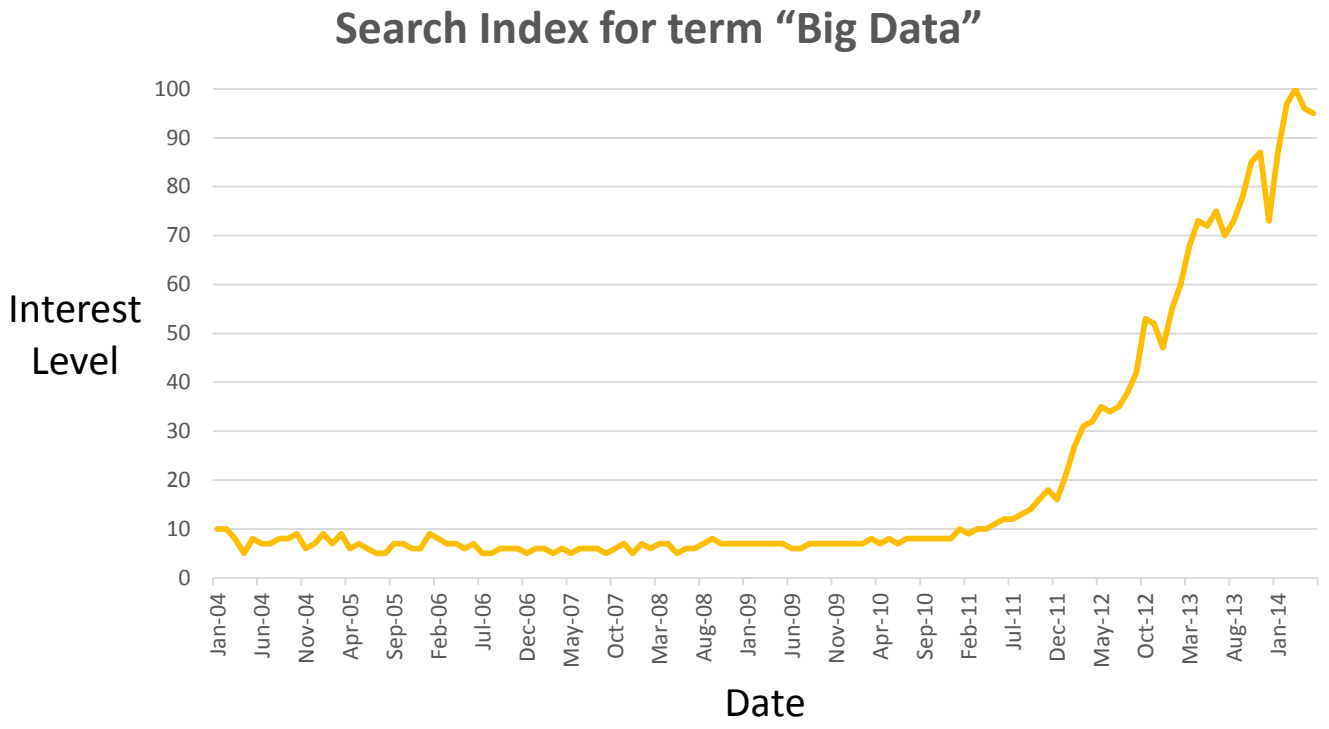
cre

BIG DATA: KEY ASPECTS

> Why Big Data?
> Structuring the Streams of Big Data
> Big Data Defined
> Traditional Market Research
> Data Quality
> Data Science Talent
> Privacy
> Getting Started with Big Data
> Marketplace Feedback

cre

WHY BIG DATA?
BIG DATA INTEREST OVER TIME

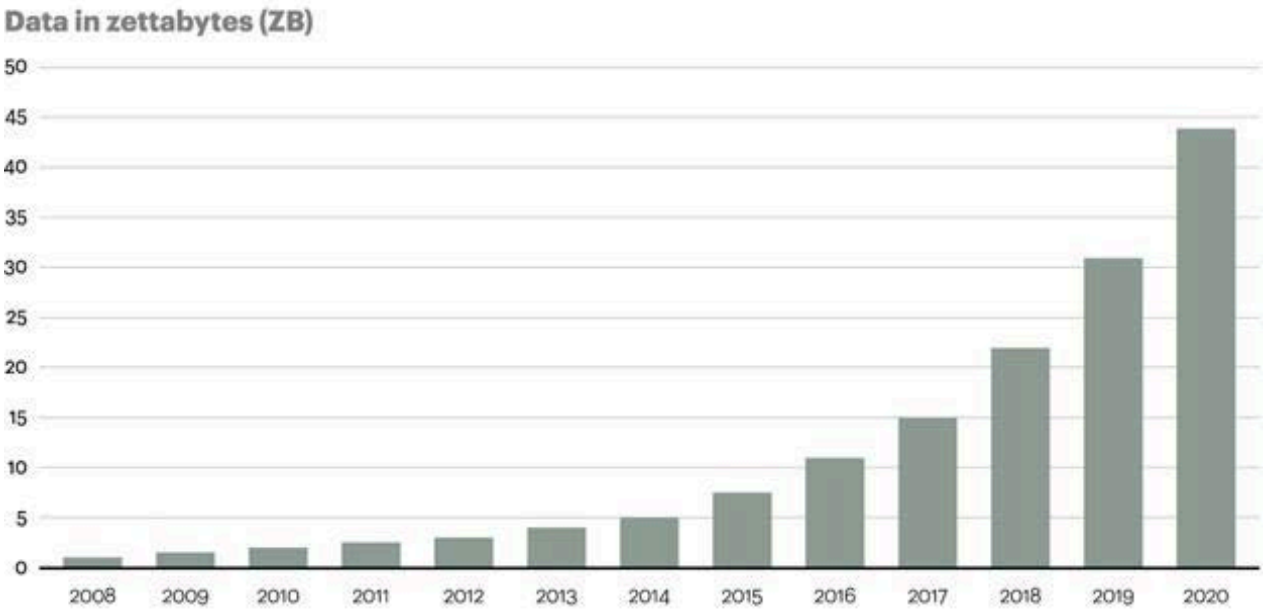


Source: Google Trends – normalize search volume according to the ratio of term(s) to the entire volume of search

WHY BIG DATA?
FUTURE GROWTH OF BIG DATA VOLUME

Figure 1

Data is growing at a 40 percent compound annual rate, reaching nearly 45 ZB by 2020



Source: Oracle, 2012

STRUCTURING BIG DATA

Data require shaping to enable comparison with structured data like TV ratings and retail sales

Social media conversations

Geo-location coordinates

Social media pictures

Mobile app usage

Retail traffic patterns

Video water marking

Audio water marking

Social media graphics

BIG DATA DEFINED: IT'S NOT JUST ABOUT SIZE

Big Data in Marketing and Advertising (BDMA):

> Too big to handle on single file server

> Most likely includes unstructured data

> Multiple data sources, reflecting consumer touch points

> Complexity depends on marketing/advertising question

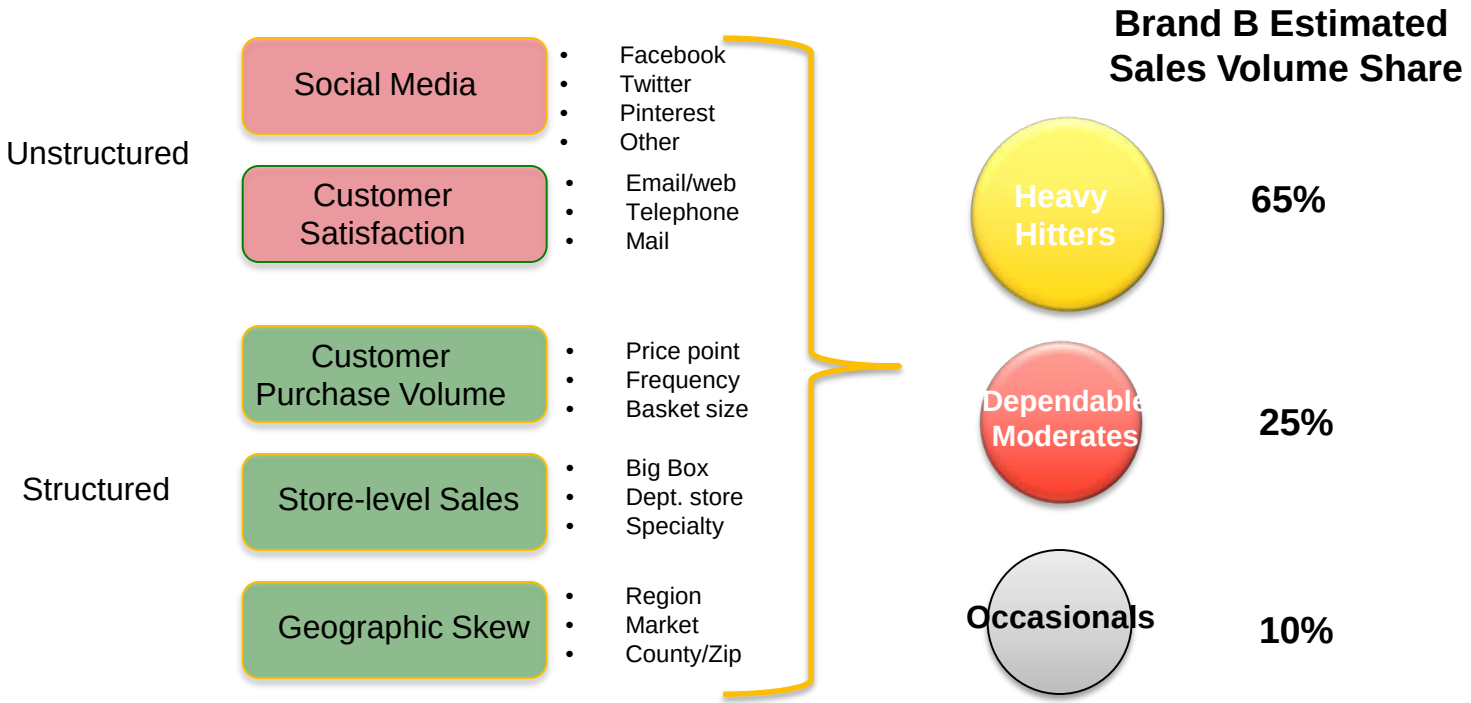
cre

BIG DATA & DECISION MAKING

> Surfaces insights and facilitates feedback immediacy not possible with traditional analytic/research approaches

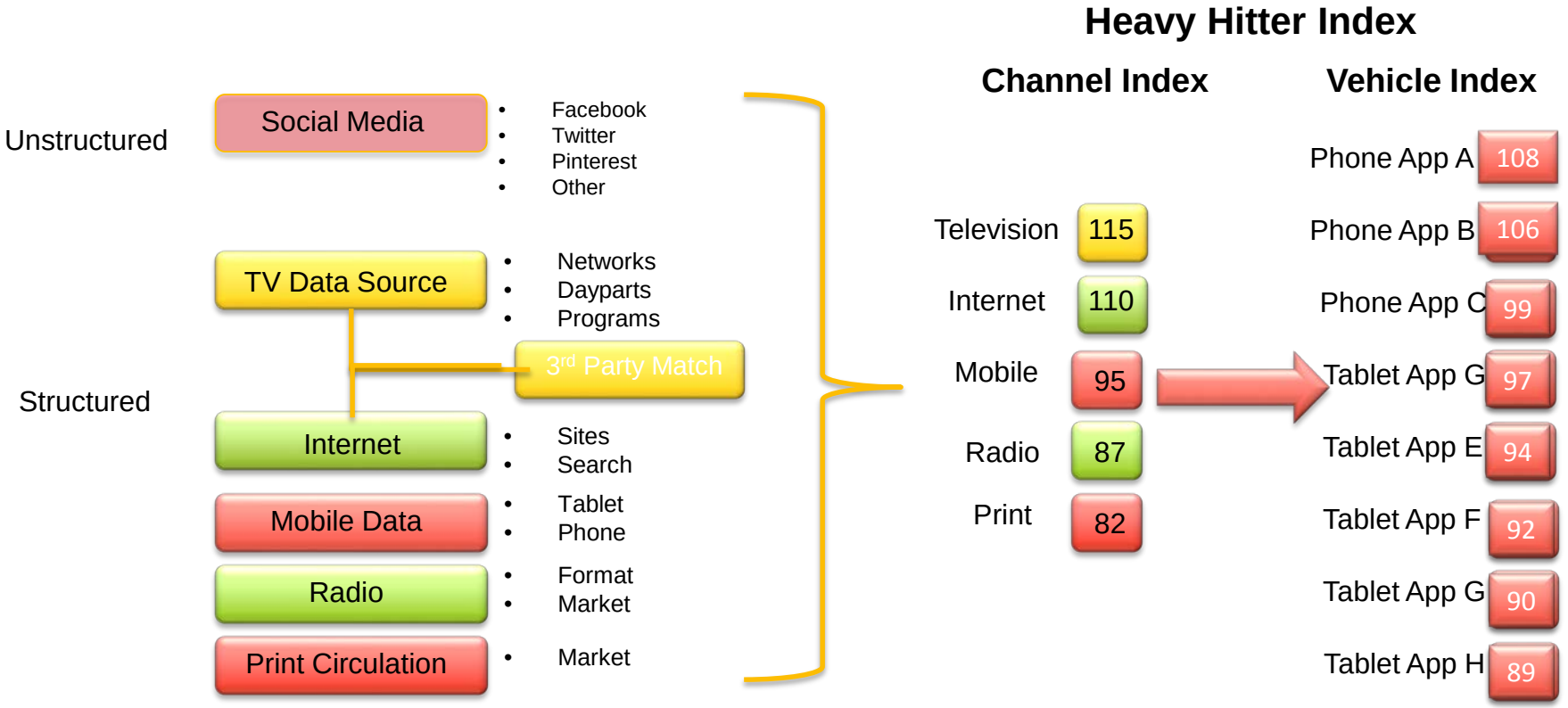
TARGETING INSIGHTS - MARKETER A, BRAND B

POOLING DATA TOGETHER TO IDENTIFY HIGH-PROPENSITY CONSUMERS, AKA THE HEAVY HITTERS

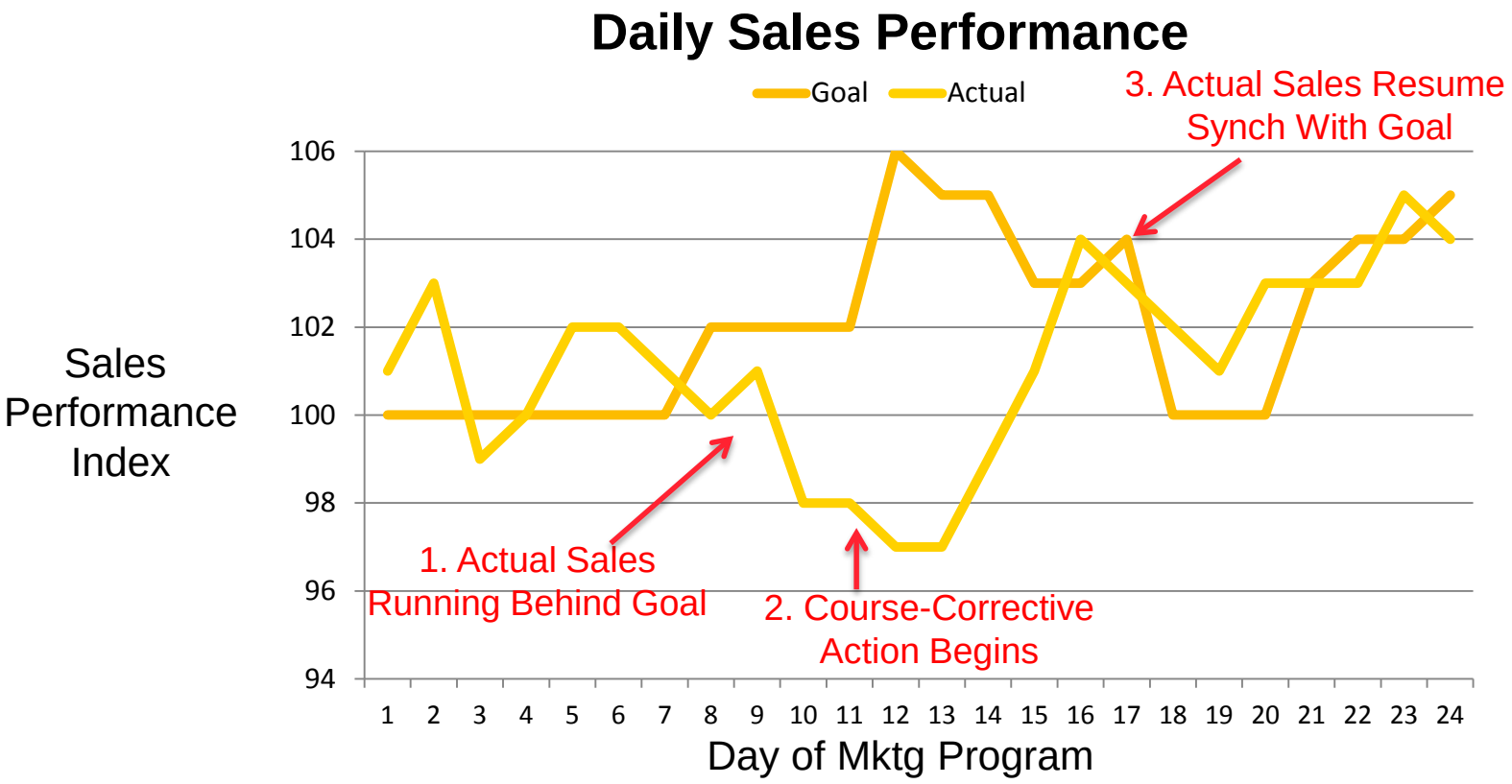


TARGETING INSIGHTS - MEDIA AGENCY

IDENTIFYING EFFECTIVE MEDIA THAT DELIVER HEAVY HITTERS . . .



FEEDBACK IMMEDIACY
ALWAYS ON, ELECTRONIC ACCESS = OPPORTUNITY FOR ADJUSTING COURSE



DATA QUALITY GOAL – CLEAN, FIT AND VALID

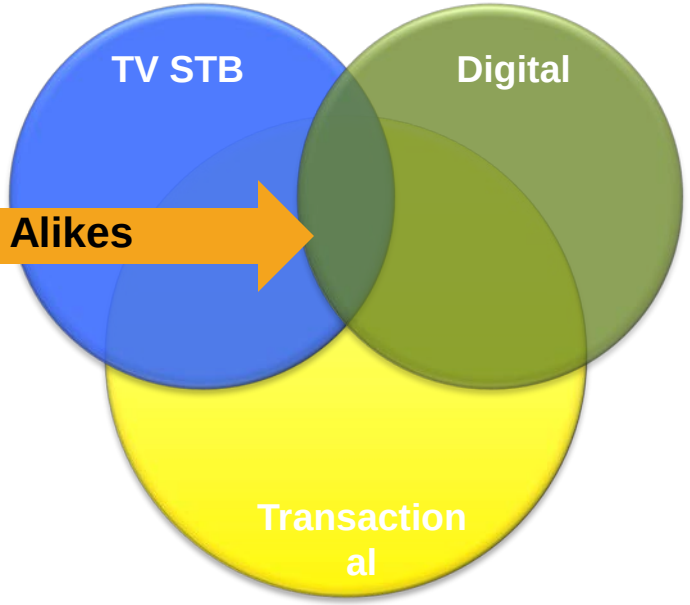
Data integrations or “mash-ups” beg questions of quality and comparability

Golf Club Bag Prospects
Customer Database

Annual Golf Occasions 12+
Annual Golf Travel 2+
Medium/Hvy Golf TV Viewer
Own specialty Golf Clubs
Pacific/Northeast Resident
Age 55+
Medium/Hvy Skier
Own High-End SUV

Find The Look Alikes

Golf Club Bag Prospects
External Databases



BIG DATA QUALITY: WHAT TO LOOK FOR

“Careful inspection of the underlying representativeness, ensuring consistency or reported metrics over time and understanding how data collection might impact accuracy.” George Ivie, Executive Director, Media Rating Council (MRC)

1. Underlying Data Values
2. Time Period
3. Representation
4. Consistency

DATA SCIENTISTS: RARE BIRDS

They’re not your traditional media or marketing research analyst

- Ability to organize/work with large data sets
- Advanced statistical background
- Recruited directly from academia or outside ad industry
- In short supply

GETTING STARTED IN BIG DATA

One part strategy, one part technology

- Tangible goal and strategy statements
  - TV network – “reduce social media post storage costs by 25%”
  - Marketer – “Uncover new target segments within customer data base and the touch points to reach them”

- Technology implementation (examples)
  - Hadoop – enables multi-server processing of large data sets
  - MapReduce – algorithmic framework within Hadoop; “air traffic controller”

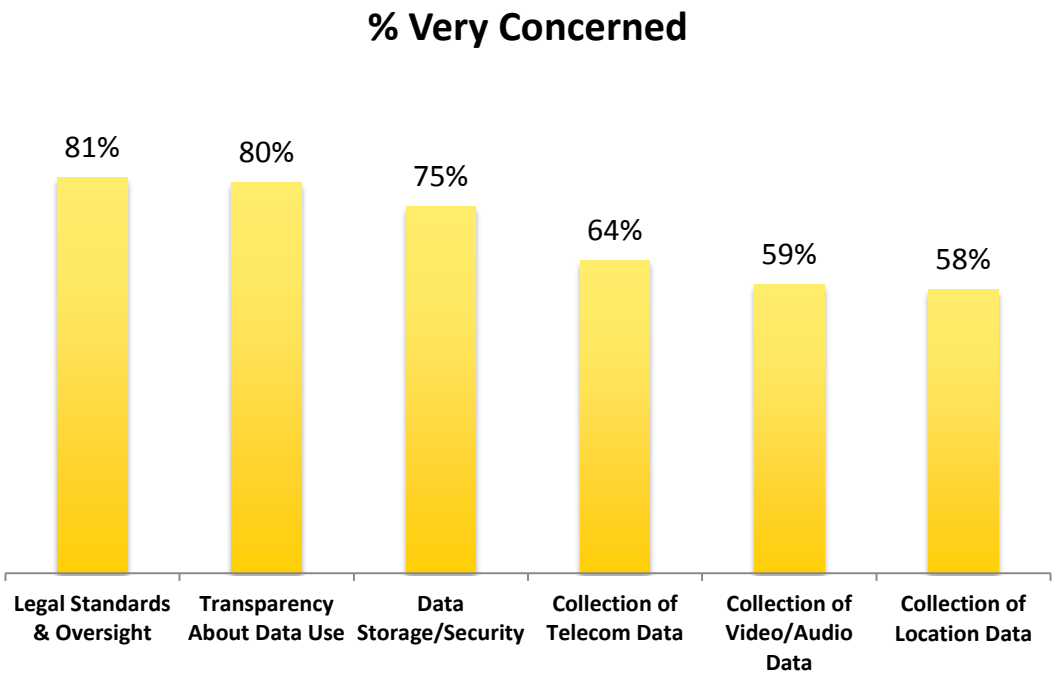
cre

TECHNOLOGIES FOR BIG DATA

Technology	Definition
Hadoop	Open-source software for processing big data across multiple parallel servers.
MapReduce	The architectural framework on which Hadoop is based
Scripting languages	Programming languages that work well with big data (e.g., Python, Pig, Hive)
Machine learning	Software for rapidly finding the model that best fits a data set
Visual analytics	Display of analytical results in visual or graphic formats
Natural language processing (NLP)	Software for analyzing text—frequencies, meanings, etc.
In-memory analytics	Processing big data in computer memory for greater speed

Source: Thomas Davenport, “big data @work,” Harvard Business Press, 2014

PRIVACY
PUBLIC CONCERN SPANS VIRTUALLY EVERY ASPECT OF BIG DATA



Source: White House, Office of the President, May 2014, 24,092 respondents

cre

MARKETPLACE FEEDBACK*

- Successful Deployment Requirements
  - strategy
  - management support
  - data talent
  - internal education
- Hottest Areas in Marketing and Advertising
  - targeting and addressability
  - creating consumer multi-touch-point profiles
  - speedy decision making
- Biggest Implementation Challenges
  - Attracting data science talent
  - integrating data from disparate sources

* Ten companies comprised of marketer, media agency, media firm, research/data firm

cre

MARKETPLACE FEEDBACK – WE'RE AT THE EARLY STAGES

"Difficult for unacquainted to understand what they can get from the data. What is the question? The push must come from the top." ***VP Sales/Analytics, MVPD***

"*Only the most advanced companies have a truly structured plan, detailed by objectives and data sources. Companies still in early adopter stage, still trying to figure out what it (Big Data) means."* ***VP of Partnership Development, Big Data Syndicator***

"Critical that it's woven into your business processes so the organization knows what to do with it." ***COO, Digital DSP***

"Biggest challenge is finding quality human resources, gathering and reporting integrated touch point data." ***SVP, Director of Analytics, Media Agency***

CLOSING REMARKS