





ANALYTICS MATURITY: INNOVATIVE APPROACH TO AUDITING DATA

STEVEN RANDALL
RAYMOND BEES

VONYA GLOBAL
APRIL 9, 2018

INTRODUCTIONS

Steven Randall

Steve has over 25 years of management and consulting experience with specialized skills in project management, operational efficiency, internal control, and conflict resolution. He provides independent counsel, identifies core issues, develops strategic solutions, and provides leadership through tenuous situations.

Steve has published many articles covering topics including leadership, risk management, internal audit, and data analytics.

Raymond Bees

Ray begin his audit career at a progressive Fortune 500 company & audit shop that included Control Self-Assessment, Risk Based Auditing, Integrated IT/Finance teams and Data Analytics in its GRC/audit "arsenal".

He has continued to expand his use and reliance on data analytics to this day leading Fortune 100/500 global audit teams and building on career experience across business/process integration, systems implementations and data governance.



A group of people are shown from the chest up, standing on a roller coaster car. They are all smiling and have their hands raised high in the air, expressing excitement. The background shows the white metal structure of the roller coaster against a clear blue sky. The text 'Internal Audit' is overlaid in the center of the image.

Internal Audit

PURE EXCITEMENT!



...and maybe a little scary!

Development & Deployment of Technology



Young Brilliant Minds

"The millennial generation, born between 1980 and 2000 now entering employment in vast numbers, will shape the world of work for years to come."



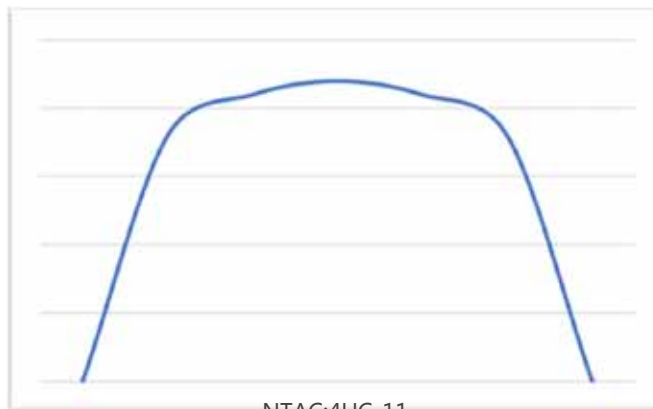
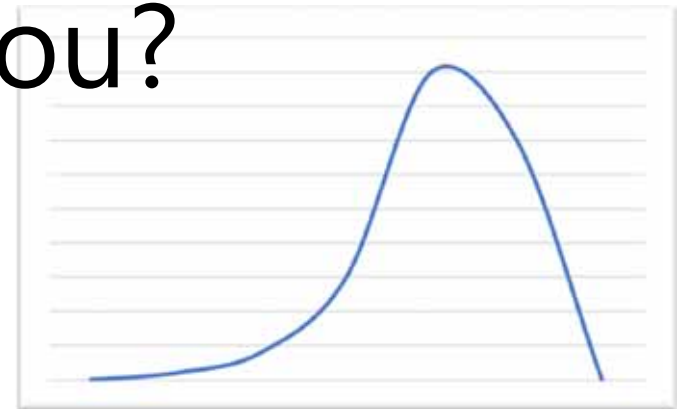
"Attracting the best of these millennial workers is critical to the future of your business. Their career aspirations, attitudes about work, and knowledge of new technologies will define the culture of the 21st century workplace."

Explosion of Data

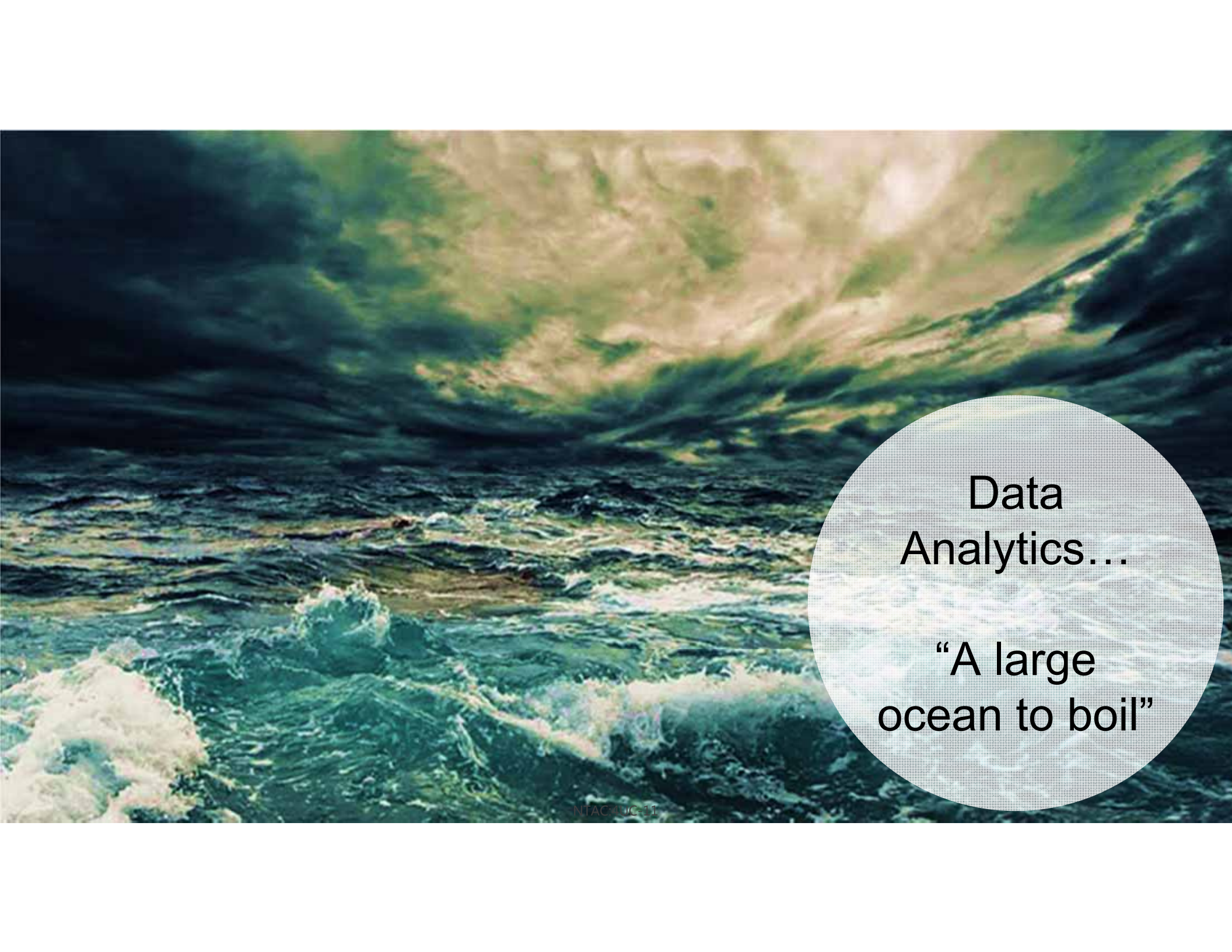


Maturity Assessment

Do I know you?



NTAC:4UC-11

An aerial photograph of a vast, turbulent ocean with deep blue and green waves. A large white circular overlay is positioned on the right side of the image, containing text. The text is centered within the circle and reads "Data Analytics..." followed by "A large ocean to boil" in quotes.

Data
Analytics...

“A large
ocean to boil”



CHICAGO AUTO SHOW



NTAC4UC-11



NTAC:4UC-11



NTAC4UC-11



NTAC:4UC-11



Objectives

A close-up photograph of a person's right hand holding a black pen, positioned as if about to write on the word "Objectives". The hand is wearing a light blue dress shirt cuff and a dark suit jacket sleeve.

1. Tips for getting started
2. Understand the types of data available
3. And how to get/source that data
4. Examples of Innovative Data Analytics



Goal: Sharing lessons we have learned to help you avoid same Pitfalls...



def·i·ni·tion
defə¹niSH(ə)n

noun

a statement of the exact meaning of a word,
especially in a dictionary.



What is Data?

Recorded information

Recorded = available at a later time

Information = available or learned facts



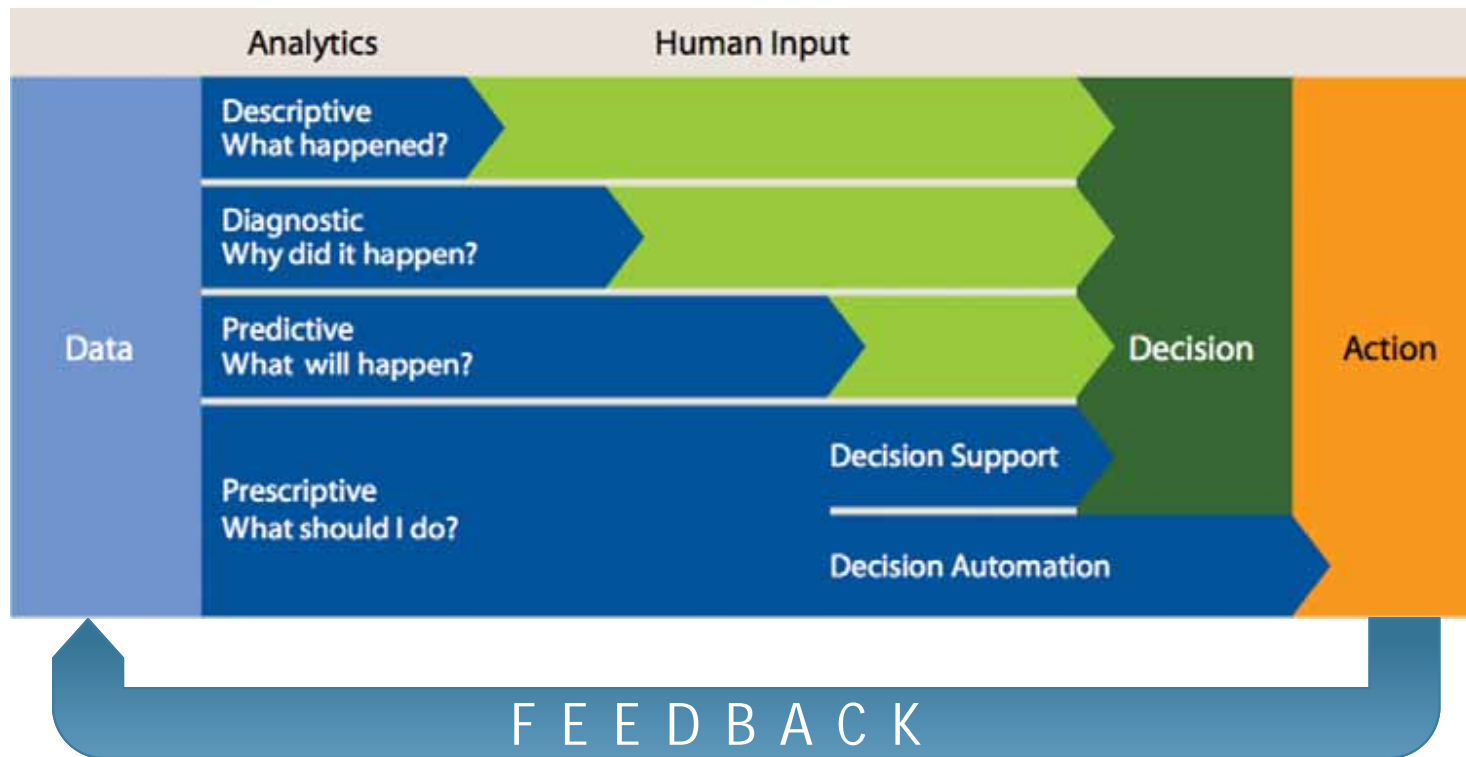
What is Data Analysis?

Process of inspecting, cleansing, transforming, and modeling
data

...with the **goal** of **discovering useful information**,
suggesting **conclusions**, and supporting **decision-making**.

NTAC:4UC-11

DATA ANALYTICS – A PROCESS MATURITY VIEW



GETTING STARTED - TOOLS

Variety of Tools

- Microsoft Excel
- Microsoft Access
- ERP queries
- Audit Tools (ACL/IDEA)
- BI Tools (Tableau)
- SaaS – Software as a Service
- Specialized Reporting
- Statistical Analysis

Features

- Users/licenses
- ERP/Data Integration
- Standard analysis
- Specialized Analysis
- Automation/Integration
- User vs. IT Tool
- Cleansing/
normalizing



NTAC:4UC-11

GETTING STARTED - HARDWARE

Need vs. Budget

- Single user?
- On the road?
- Ad hoc vs. recurring?
- Size of data sets?

Features

- Laptop/desktop/network
- Processor
- Memory
- Storage



IIA CHICAGO CHAPT

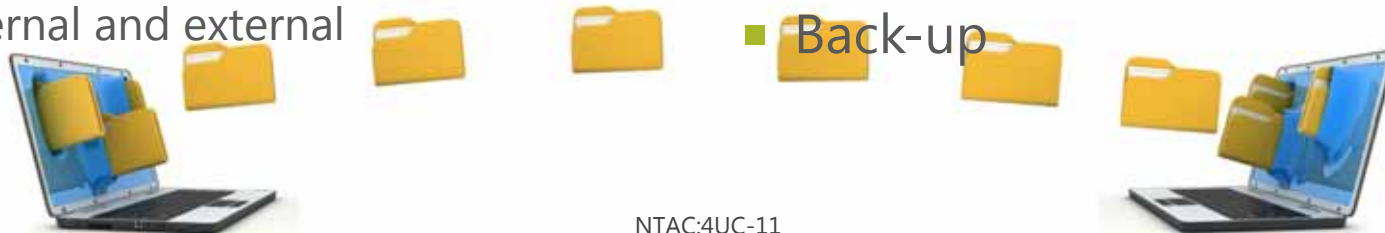
GETTING STARTED – TRANSFERRING / STORING DATA

Considerations

- Size of your data (multiply it)
- Frequency of downloads
- Where is your data
- Where are you?
- Transfers
 - local and remote
 - internal and external

Resources

- USB/thumb drives
- External drives
- CD/DVD
- Network/shared drives
- Internal share sites
- External share sites
- Back-up



GETTING STARTED – INTERNAL RESOURCES

People

- Knows the data
- Knows IT
- Knows the business
- Access to the data
- Scrubs the data
- Normalizes the data
- Tactical Tech user
- Knows the DA software
- Knows auditing

Documents

- System Architecture
- Data layouts/DLL
- Table layouts
- Interfaces
- Process/data flows
- Operating Manuals



GETTING STARTED – PROJECT OBJECTIVES

Set reasonable goals...start small/dream big

- Limit the scope (time period, process area, multiple systems & data sets)
- Use available data
- Limit joins & logic
- Understand the data before analyzing it
- Test your analysis & assumptions to real examples
- Don't forget data integrity controls

NTAC:4UC-11



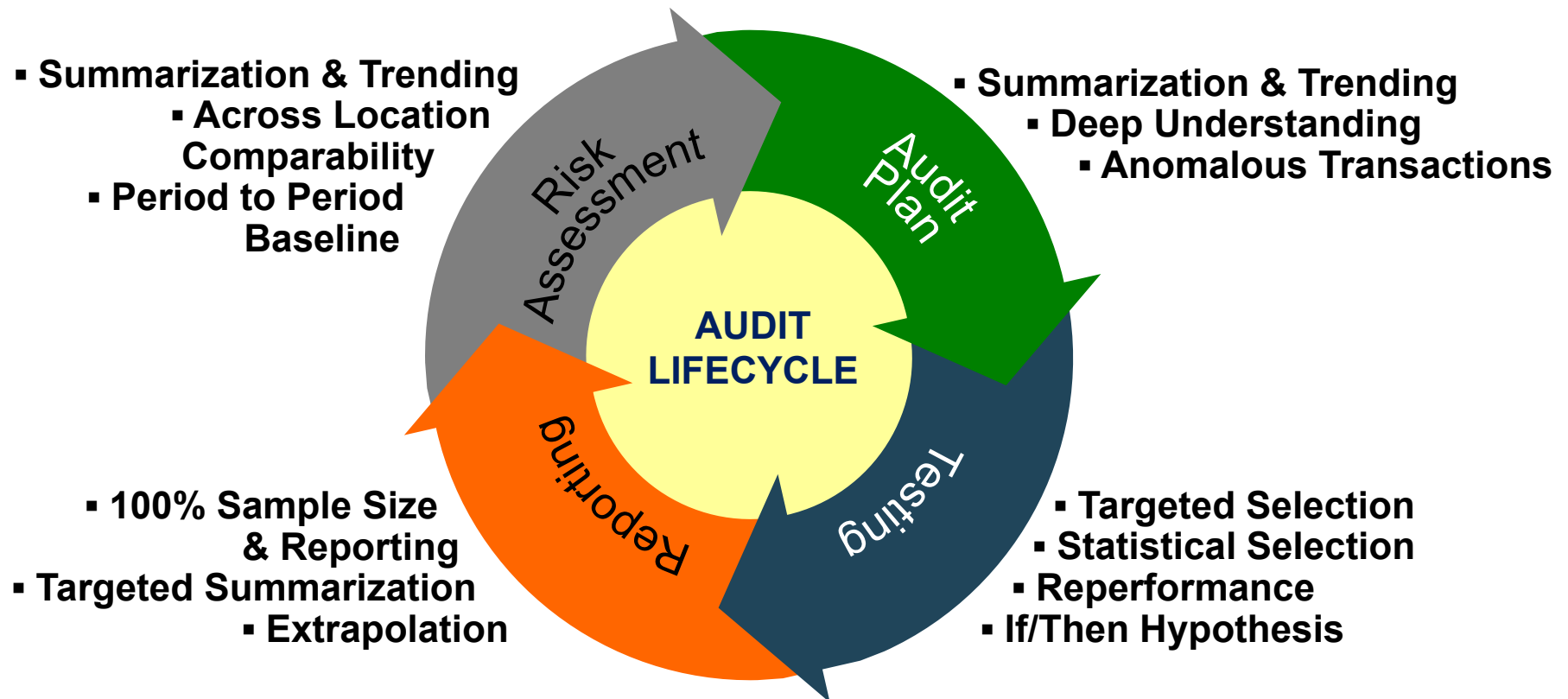
GETTING STARTED – PLANNING CONSIDERATIONS

Design your analysis around the data. You may need to modify or curtail your initial goals...

You are constrained by multiple factors:


- the process(es) that created the data
- your knowledge of the data
- data available
- the frequency of its collection
- the volume/periods retained
- the level of detail/data increment
- data quality

DATA ANALYTICS USE ACROSS THE AUDIT LIFECYCLE



Role of Internal Audit and Data Analysis

Internal Audit

A large yellow circle containing a list of internal audit functions.

- Financial Reporting (assertions)
- Control Design and Operation (ICFR & Operating Controls)
- Compliance (external + internal)
- Operating Effectiveness & Efficiency
- Understanding/Information Gathering
- Assessment/Validation
- 3rd Party Performance

Role of Internal Audit and Data Analysis

Internal Audit

Data Analysis



**Data Analysis likely has a role –
regardless of the nature of your audit...**

- Financial Reporting (assertions)
- Control Design & Operation (CFR & Operating Controls)
- Compliance (external + internal)
- Operating Effectiveness & Efficiency
- Understanding/Information Gathering
- Assessment/Validation
- 3rd Party Performance

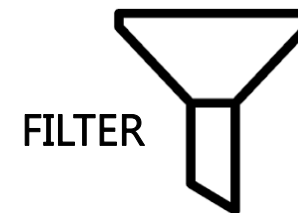
USES & BENEFITS OF DATA ANALYSIS

Discovery
Understanding/Summarization
Targeted Identification
Causation/Extrapolation

SUMMARIZE



NTAC:4UC-11





Data is like any other language...

"Data represents events and is a function of the system and processes that created it, its use may evolve and change over time and in some instances it may be more precise than in others, where different users may have different syntax, uses and interpretation."

It is important to understand the context of its use and how it was created to properly analyze the data and draw conclusions.

WORKING WITH DATA – DATA TYPES (BY USE)

- Data settings/parameters – the configuration that drives how a system works
- Master Data – rules for grouping transactional data, recurring values
- Transactional Data – the lowest level of data collection
- Summary Data – aggregated data (account balances, document totals)
- Audit trails and logs - witness to history
- Meta Data – understanding the operation and relationship of data thru data governance that defines and enforces those relationships

How many shortest-length paths are there to get from your house to the doughnut shop?



$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

$$e^{i\pi} + 1 = 0$$

P	Q	R	PVQ	PVR	(PVQ) ^ (PVR)
T	T	T	T	T	T
T	T	F	T	T	T
T	F	T	T	T	T
T	F	F	T	T	T
F	T	T	T	T	T
F	T	F	T	F	F
F	F	T	F	T	F
F	F	F	F	F	F

7, 11, 15, 19, 23...

$$A_1 - A_0 = 4$$

$$A_2 - A_1 = 4$$

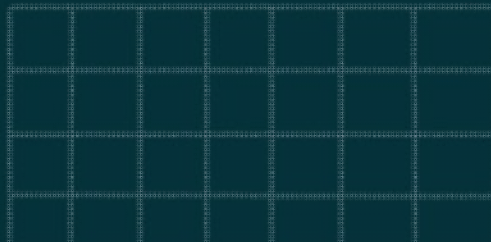
$$A_3 - A_2 = 4$$

$$\vdots$$

$$A_n - A_{n-1} = 4$$

$$A_n = A_{n-1} + 4$$

4 up's
↑
7 right's
→



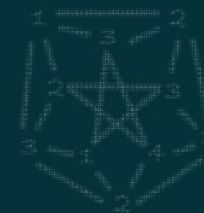
101

101

101000101



WORKING WITH DATA – DATA STRUCTURE



Converse:
 $\exists x \forall y (x > y + 1 \rightarrow x \geq 2y)$

Negation:
 $\neg [\exists x \forall y (\neg (x \geq 2y) \vee x > y + 1)]$
 $\forall x \exists y (x \geq 2y \wedge x \leq y + 1)$

Contrapositive:
 $\exists x \forall y (x \leq y + 1 \rightarrow x < 2y)$

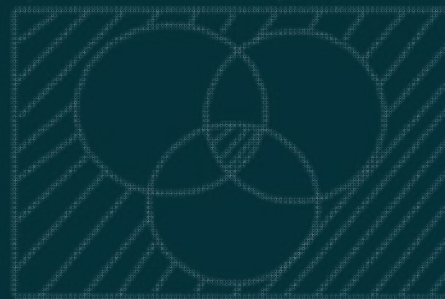
$$v = e + f = 2$$

P.I.E. Example:

$$6! - \left[\binom{6}{1}5! - \binom{6}{2}4! + \binom{6}{3}3! - \binom{6}{4}2! + \binom{6}{5}1! \right]$$

There are six dogs to give 13 tacos. use a 'stars and bars' diagram to illustrate the first and sixth dog get 3 tacos, the second dog gets none, the third dog gets 5 and the fourth dog gets one.

$$(A \cup B \cup C) \cup (A \cap B \cap C)$$



NTAC4UC-11

☆☆☆||☆☆☆☆☆|☆||☆☆☆|

$$A = \{2, 4, \text{Ⓢ}, \text{Ⓢ}\}$$

How many shortest-length paths are there to get from your house to the doughnut shop?



$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

P	Q	R	PVQ	PVR	(PVQ) ^ (PVR)
T	T	T	T	T	T
T	T	F	T	T	T

7, 11, 15, 19, 23...

$$A_2 - A_1 = 4$$

WORKING WITH DATA – DATA STRUCTURE

Structured Data

Defined data types, and/or range of values

- Structured data may be less structured – than expected.
- Consider data governance – defining/enforcing standardization;
- Input controls, validation/exception reporting & process maturity

Unstructured Data

Such as text/narrative and free entry fields

- This data can provide context, though it is not usually the first source of data for analysis
- Not consistent, or always available

WORKING WITH DATA – LAYERS OF DATA

Data may exist at multiple levels at varying levels of detail

Example – sales of a movie title (5 layers to determine title)

1. A DVD can contain multiple titles
2. A DVD set has multiple DVD discs
3. A DVD collection has multiple DVD sets
4. DVDs are distributed in pre-configured/bundled cartons mixing different sets/collections
5. DVD cartons are sold in pre-configured pallets of cartons

WORKING WITH DATA – EXPLODED DATA

PALLET					CARTON					BOX SET						
PROD #	DESCR.	QTY	UNIT COST	EXT. COST	PROD #	DESCR.	QTY	EXT. QTY	UNIT COST	EXT. COST	PROD #	DESCR.	QTY	EXT. QTY	UNIT COST	EXT. COST
PW52NR	Rel. W52	2	\$xx,xxx.xx	\$xx,xxx.xx												
					C2750	Hitchcock	10	20	\$xxx.xx	\$x,xxx.xx						
					C1892	Sci-Fi	12	24	\$xxx.xx	\$x,xxx.xx						
					C0978	Romance	20	40	\$xxx.xx	\$x,xxx.xx						
					C7901	NR W50	23	46	\$xxx.xx	\$x,xxx.xx						
					C3754	NR W52	21	42	\$xxx.xx	\$x,xxx.xx						
					C2232	DreamWrk	14	28	\$xxx.xx	\$x,xxx.xx						
											VER012	Vertigo Dir.	2	40	\$xx.xx	\$xxx.xx
											VER001	Vertigo	4	80	\$xx.xx	\$xxx.xx
											ROP001	Rope	2	40	\$xx.xx	\$xxx.xx
											NNW007	N by NW	6	120	\$xx.xx	\$xxx.xx
											PSY001	Psycho	4	80	\$xx.xx	\$xxx.xx
											BIR012	The Birds	4	80	\$xx.xx	\$xxx.xx
											TOP001	Topaz	2	40	\$xx.xx	\$xxx.xx
											MAR001	Marnie	3	60	\$xx.xx	\$xxx.xx
											THI012	Catch Thief	7	140	\$xx.xx	\$xxx.xx
											COM001	Complete	1	20	\$xx.xx	\$x,xxx.xx
											SAB003	Saboteur	3	60	\$xx.xx	\$xxx.xx
											DIA011	Dial M	2	40	\$xx.xx	\$xxx.xx

At level 3 with 2 more levels to go...

1 row → 19 rows/records

5 cells → 323 cells/fields (5 → 17 wide)

WORKING WITH DATA – DATA LAYERS/VARIANTS

Blade Runner

- 1982 or 2017 version?
- Blu-ray or DVD?
- Theatrical Version?
- Un-rated version?
- Directors cut/Behind the scenes?
- With out-takes?
- Bonus set?
- Anniversary edition?
- Television version/Airline version?
- French/Spanish/English?
- English (US, Canada, UK or Int'l)?

Packaging Variants

- Bonus coffer
- Stand-up display
- Limited edition with book
- Plastic case
- Case with security tag
- Over-wrap in English/French
- Box set with discount coupon

WORKING WITH DATA – SOURCING DATA

Moral of the story...Understand your data, make smart choices to allow insightful analysis

Know your apples

Know your oranges



No fruit salad

Most analysis errors stem from a lack of understanding of the data ... leading to problems with how it's used, and interpreted.

WORKING WITH DATA – SOURCING DATA

How to get workable data...

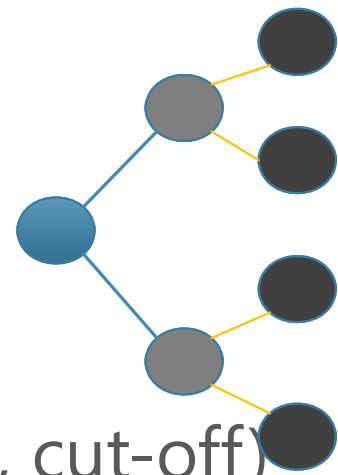
- Source centrally where possible (for consistent treatment)
- Source from existing reports or summary tables/views
- Limit joins and introduced logic to combine information
- Start with a workable size data set for proof of concept
- Consider changes/variations over time when testing
- Validate to source data



WORKING WITH DATA – DATA GENERATIONS

The impact of different generations of data

- Different uses and definitions
- New/different, layered-on information
- Different retention
- Different controls/lack of controls (i.e. quality, cut-off)



Different needs: The ultimate reason for different uses/systems and the risk when using

WORKING WITH DATA – CLEANSING DATA

You may need to cleanse your data for inherent errors – to be able to draw accurate/consistent conclusions

- Be prudent in your use of cleansing and the logic employed
- Evaluate source data errors vs. sourcing errors
- Document the changes made (in case they need to be reversed/modified)
- Test your cleansed data



WORKING WITH DATA – PROXIES

Example – Cost Data

- Cost vs. Standard Cost + Variance
- Actual vs. Estimate
- Summarized Data
- Approximations & Allocations

Sources of Information

- Project Estimates & Budgets
- Purchase Orders (POs)
- Asset Ledgers
- General Ledger
- Accounts Payable
- Disbursements Registers

WORKING WITH DATA – WHEN YOU THINK YOU’VE FAILED



You haven't failed – you have a different story to tell...

- Your analysis has revealed other issues to be considered:
 - Data quality/integrity issues (data governance + controls)
 - Adequacy of audit trails
 - Data aggregation and sufficiency of detail
 - Data retention limitations
 - Other...

WORKING WITH DATA – (NO) HARD CONCLUSIONS

Black & White vs Shades of Grey

- There are no hard, indisputable conclusions from data analysis
- Data represents activity, using defined processes
- While a field may contain or not contain a value – that alone is not conclusive, the value/lack of value must be interpreted relative to process requirements and norms



WORKING WITH DATA – CONTROL ASSESSMENT

Operation of Control vs. Quality of
Operation

Through data analysis you may determine if a control was performed, but it is more difficult to determine if it was performed properly (i.e. the quality of performance)

WORKING WITH DATA – INFERENCES

Data analysis allows you to make inferences, and propose operating models and if/then causation.

However as noted – these are not hard and fast conclusions.

You likely will need to go back to source data/documents to understand how the data represents real life.

DATA ANALYTICS EXAMPLE – G/L

Activities

- Buying
- Receiving
- Capitalizing
- Expensing
- Producing
- Consuming
- Revaluing
- Reporting

The G/L is the accounting shorthand of everything occurring at the organization

Sources

- Purchasing
- Accounts Payable
- Fixed Assets
- Payroll
- Invoicing
- Accounts Receivable
- Inventory
- Manufacturing

DATA ANALYTICS EXAMPLE – G/L DATA

- DR/CR indicator
- Account Number/Name
- Transaction amount
- Transaction currency
- JE Number
- JE description*
- Line #
- Line description*
- Transaction date
- Effective date
- Reporting period
- Source system
- Transaction Code
- JE type
- JE poster
- JE approver

* Unstructured data

DATA ANALYTICS EXAMPLE – G/L RISKS

Assertions

- Classification
- Cut-off
- Existence/Occurrence
- Valuation
- Authorization

Activities

- Management override
- Cover-up
- Segregation of Duties

Risk Indicators

- Unusual reversals
- Permanent JEs
- Timing of entries
- Retro adjustments

DATA ANALYTICS EXAMPLE – PAYROLL DATA

Master Data

- Employee name/#
- Department
- Pay rate
- Benefits
- Employee address
- EE bank account(s)
- Bank account type
- Tax rate

Transactional Data

- Hours worked
- Pay amount
- Transaction value
- Payment type
- Withholding type
- Effective date
- Transaction date
- Pay period



DATA ANALYTICS

EXAMPLE – PAYROLL

ACTIVITIES

Activities

- Pay employees (pay/overtime/commissions/bonus)
- Reimburse employees (expense reports)
- Make payments for/on behalf employees
- Make employer payments benefitting employees
- Report payroll & employee income taxes

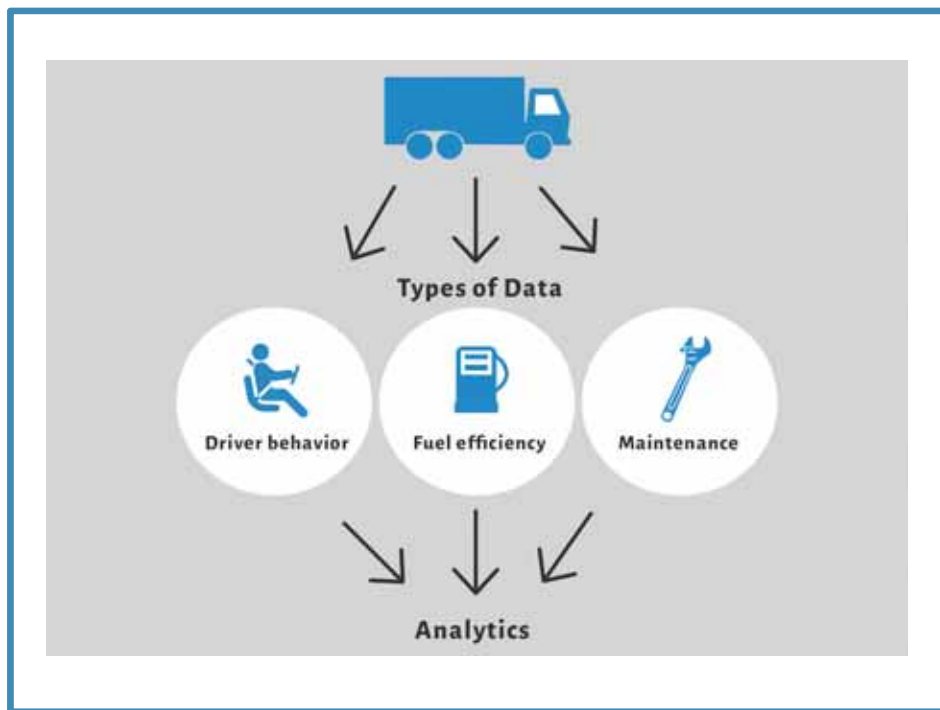
DATA ANALYTICS EXAMPLE – PAYROLL

Risks

- Ghost employees
- Skimming of pay
- Diversion of pay
- Inflated pay or hours
- Unauthorized overtime
- Unauthorized changes



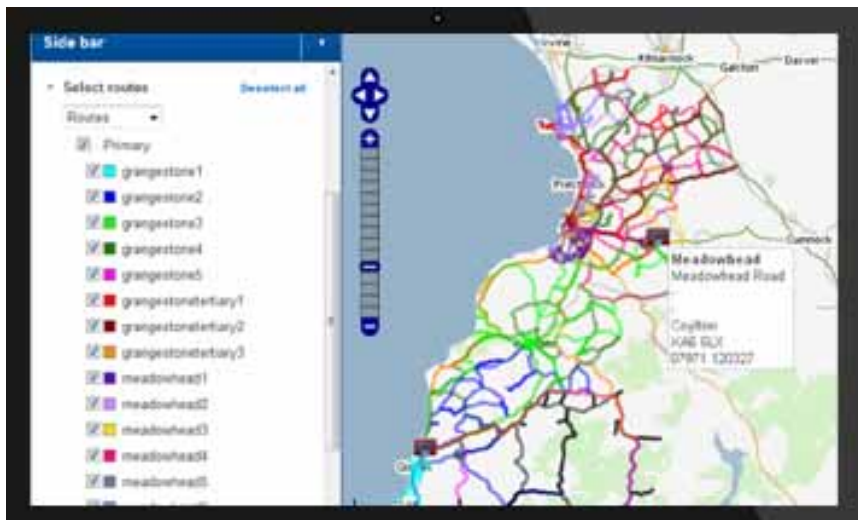
DATA ANALYTICS EXAMPLE – FLEET MANAGEMENT



Devices are embedded within vehicles to capture data

- Breaking
- Acceleration
- Driving speed
- Fuel consumption
- Driving time
- Idle time
- Location/GPS tracking

DATA ANALYTICS EXAMPLE – FLEET MANAGEMENT



Corporate Data exists for fleet management

- Route management
- Expense reporting
- Cost Analysis
- Asset Utilization
- Driver Performance
- Maintenance History

DATA ANALYTICS EXAMPLE – FLEET MANAGEMENT



DATA ANALYTICS EXAMPLE – FLEET MANAGEMENT

- Opportunity to identify misuse of vehicles
- Opportunity to identify cost savings in fuel efficiency
- Opportunity to identify cost savings in branded vs generic fueling stations
- Opportunity to identify driving characteristics that increase wear and tear on vehicle
- Opportunity to identify the driving characteristics indicative/predictive of incidents
- Supply chain optimization, inventory allocation and planning, fleet reliability



RISK ASSESSMENT EXAMPLE – ACCOUNTS PAYABLE

Trend/Compare

- Understanding of Key Activities
- Create Baseline
 - Compare Across Locations
 - Monitor Location Over Time



Transactions

- Core / Frequent
- Unusual
 - Corrections
 - Errors
 - Inefficiencies



Measure



- Volume
- Frequency
- \$ Impact
- Range (min/max) & Distribution

**THE FUTURE
IS HERE...**



NTAC/UC 11

...AS OF YESTERDAY.



Physical Inventory
Visual Data
Image Recognition
Specialized Sensors

Site Assessment
Damage Evaluation
Monitoring of Utilization



THE FUTURE IS HERE... Royalty Enforcement

- Audio / Visual Data
- Pattern Recognition
- Who's using my music?
- Where did my ad run?



THE FUTURE IS HERE...



Document Analysis

- OCR - Optical Character Recognition
- Key word search
- Key word summarization
- Contract analysis i.e. Leases (ASC 842 / IFRS 16)
- Legal eDiscover

REAL-TIME ANALYSIS

E Commerce

- Targeted Advertising
- Fraud Assessment



Cyclical Audits → Continuous Auditing → Real-Time Auditing

- Increasing frequency of data points
- Reducing data latency
- Shrinking measurement intervals



Continue the conversation
Virtual Roundtable on Data Analytics

Thursday, April 12 @ 4:00

Follow

[#InternalAuditorsRock](#)

QUESTIONS AND ANSWERS?

If you have any questions or comments please do not hesitate to contact us...

Steven Randall

srandall@vonyaglobal.com

Raymond Bees

rbees@vonyaglobal.com

THANK YOU FOR YOUR TIME AND ATTENTION!

IIA CHAPTER CHICAGO | 58TH ANNUAL SEMINAR

