

Selling Data Investments: Gaining Business & Executive Support for Data Management



Peter Aiken, Ph.D.

- I've been doing this a long time
- My work is recognized as useful
- Associate Professor of IS (vcu.edu)
- Institute for Defense Analyses (ida.org)
- DAMA International (dama.org)
- MIT CDO Society (iscdo.org)
- Anything Awesome (anythingawesome.com)
- Experienced w/ 500+ data management practices worldwide
- Multi-year immersions
 - US DoD (DISA/Army/Marines/DLA)
 - Nokia
 - Deutsche Bank
 - Wells Fargo
 - Walmart
 - HUD ...
- 12 books and dozens of articles



Program overview

Selling Data Investments: Gaining Business & Executive Support for Data Management

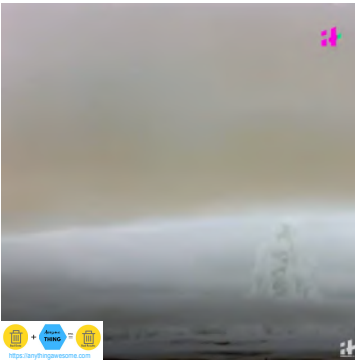
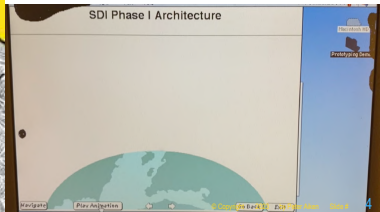
- Telling a Story (Star Wars)

- Common Vocabulary = Smooth Operations
- Objective Treatment of Governed Assets
- Data Ownership Challenges
- Data Ethics Tales
- Increasing Employee Data Literacy (Knowledge Workers)
- Productivity improvements (ChemPhDs)
- Covid undercounting
- Effective communication (Seven Sisters)
- Cultural prerequisites (Chocolate Story)

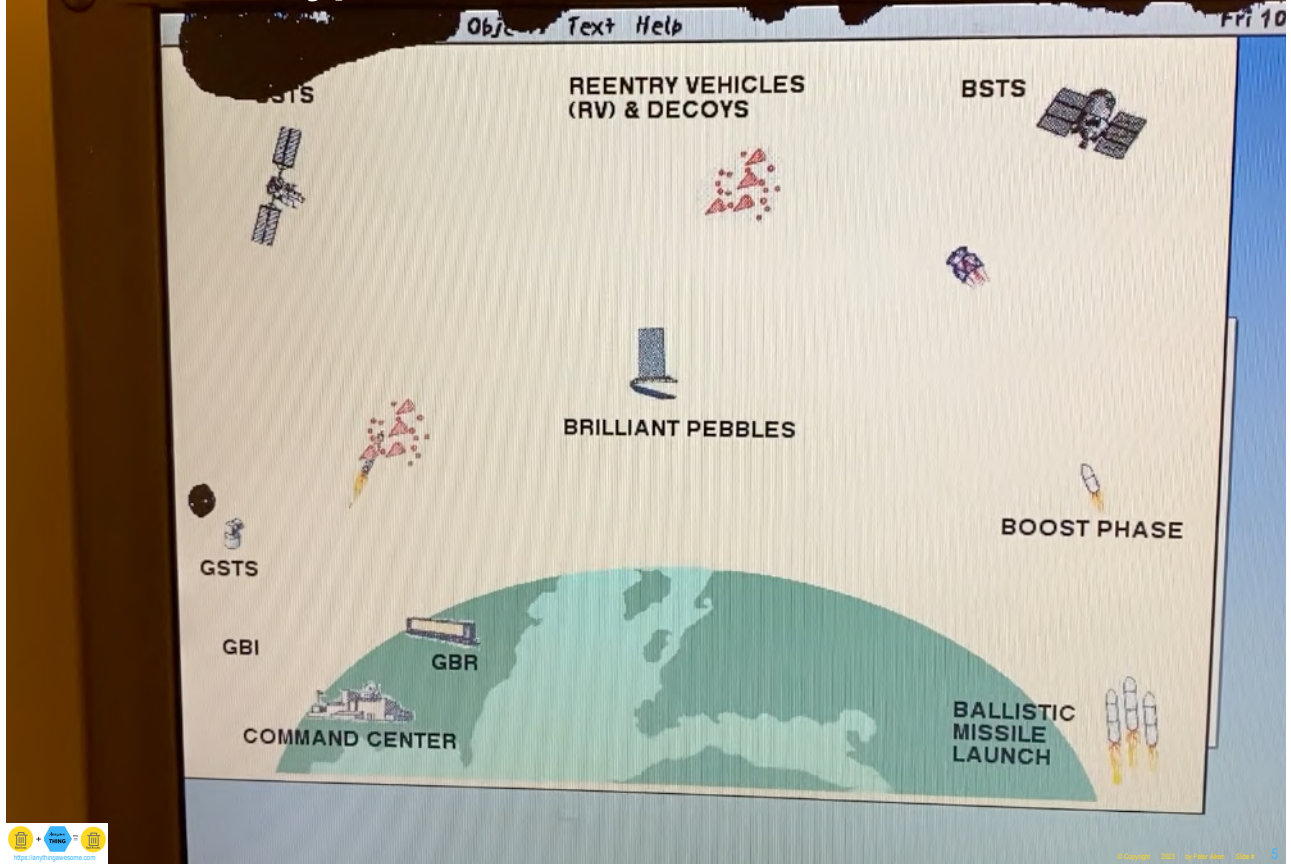


Edward Teller & Strategic Defense Initiative





SDI Prototype

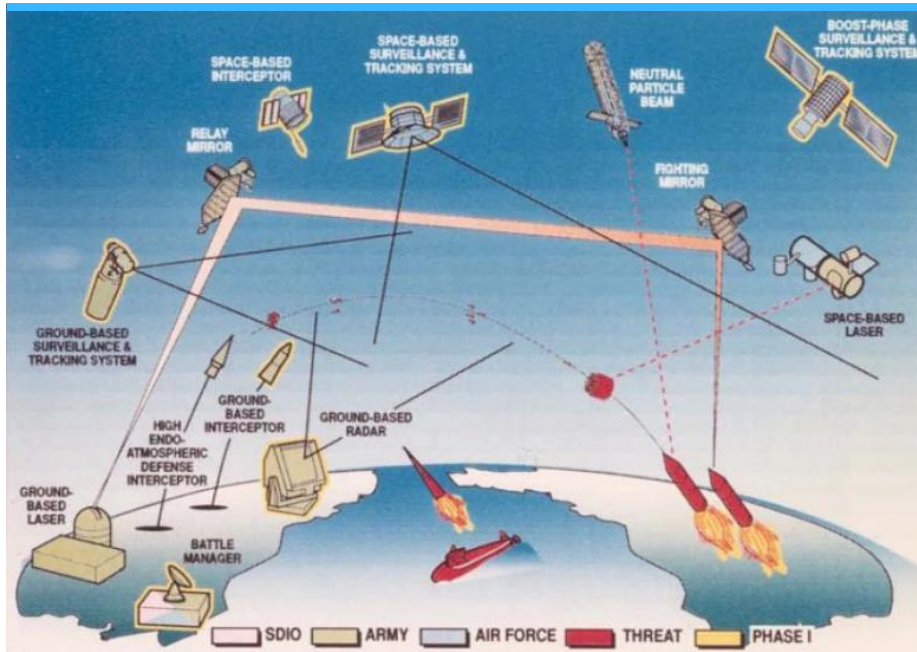


Two minute description of Star Wars

X-Ray Lasers

They are literally throwing rocks at rockets!

<https://www.youtube.com/watch?v=sMfmVzHZvkc&t=41s>



BMDS - THE BALLISTIC MISSILE DEFENSE SYSTEM

When space-based sensors and ground-based radar gather information on the target, they will transmit it to the so-called kill vehicle and to command centers.

② Kill vehicle maneuvers itself into a collision course with incoming warhead, which it will hit at about 15,000 mph at an altitude of about 140 miles.

SENSORS

Boost/Ascent Defense Segment

Midcourse Defense Segment

Terminal Defense Segment

Heat-detecting satellite sensor

Incoming missile

Intercept

Kill vehicle

Booster separates

Interceptor missile

COMMAND CENTERS

U.S.

FT. GREELY

MINUTEMAN AIR FORCE BASE

U.S.

Pacific Ocean

CHINA

JAPAN

RUSSIA

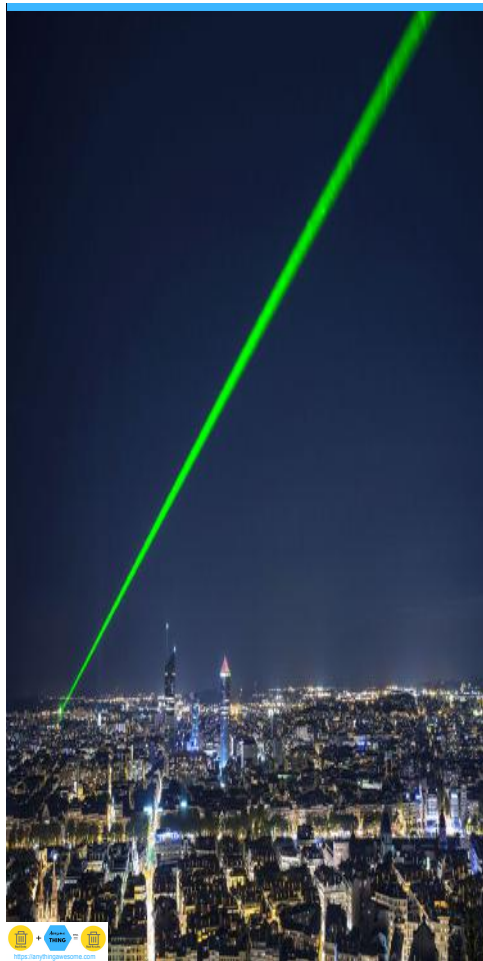
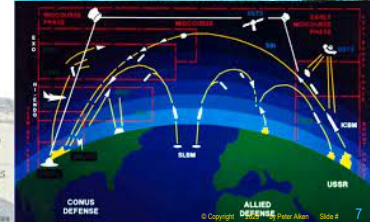
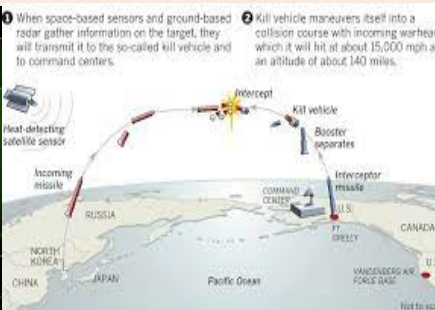
CANADA

U.S.

kill to sea

STRATCOM USMIRVCOM USPAFACM EUCCOM COMCOM

https://myingw.com



Patriot Missile

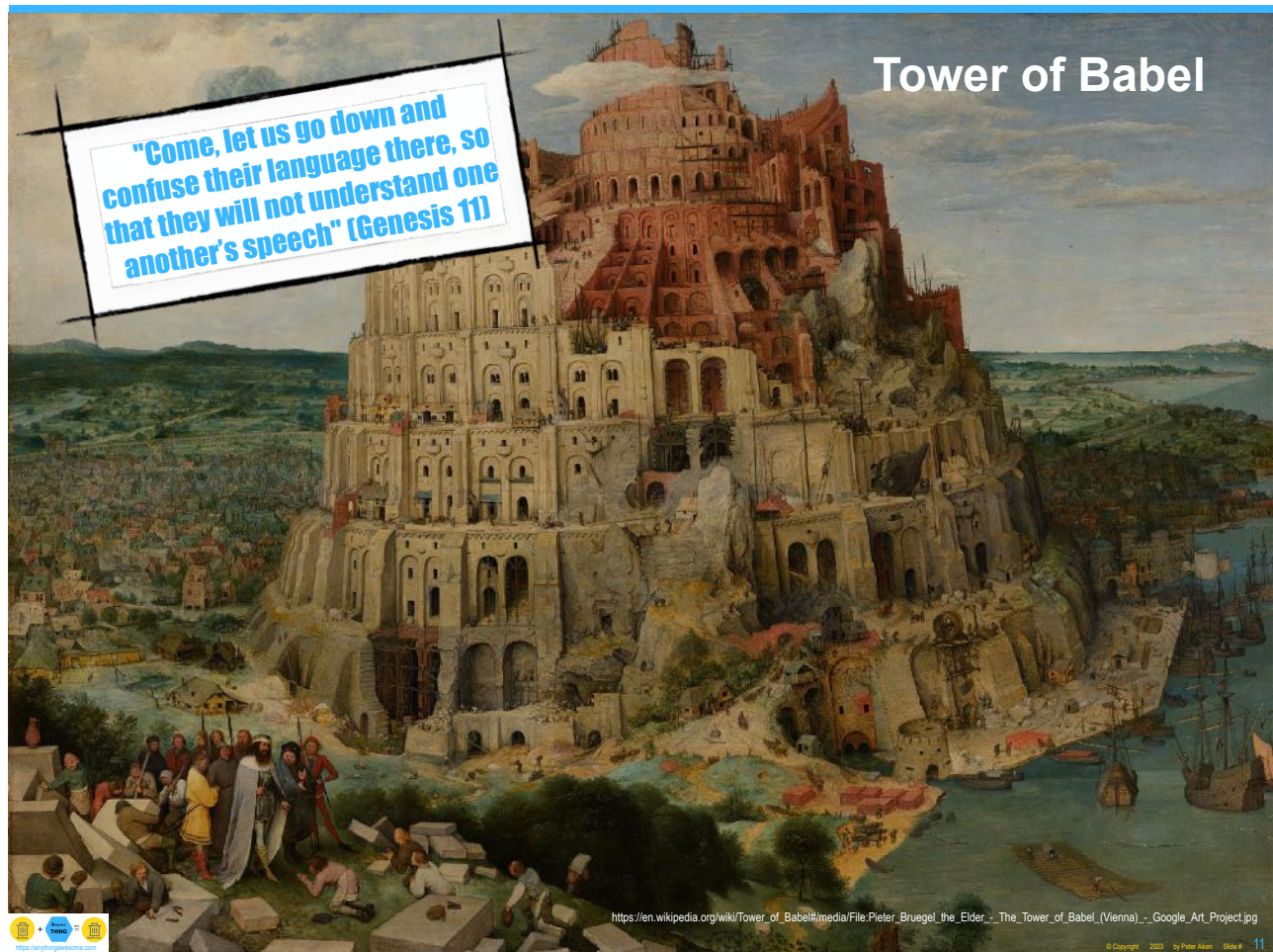


Program Overview

- Telling a Story (Star Wars)
- Common Vocabulary = Smooth Operations
- Objective Treatment of Governed Assets
- Data Ownership Challenges
- Data Ethics Tales
- Increasing Employee Data Literacy (Knowledge Workers)
- Productivity improvements (ChemPhDs)
- Covid undercounting
- Effective communication (Seven Sisters)
- Cultural prerequisites (Chocolate Story)

*Selling Data Investments:
Gaining Business &
Executive Support for
Data Management*





Tower of Babel

Data Integration/Exchange Challenges



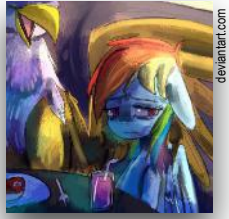
- **Customer** typically has had different meanings to different parts of the organization:
 - Accounting → organization that buys products or services
 - Service → client
 - Sales → prospect
- Assigning the same mission to the DoD ‘lines of business’ to: “Secure the building” elicits very different results from each ‘line of business’:
 - **Army**: Posts guards at all entrances and ensures no unauthorized access
 - **Navy**: Turns out all the lights, locks up, and leaves
 - **Marines**: Sends in a company to clear the building room-by-room; forms perimeter defense around the building
 - **Air Force**: Signs three year lease with option to buy [2nd example courtesy of Burt Parker]



Understanding = Interoperability



- All organizations have architectures
 - Business
 - Process
 - Systems
 - Security
 - Technical
 - Data/Information
- Some are better understood and documented (and therefore more useful to the organization)
- 'Understanding an architecture'
 - Documented and articulated as a (digital) blueprint illustrating the commonalities and interconnections among the architectural components
- Ideally the understanding is shared by
 - Business
 - Technical
 - Systems



Common vocabulary expressing integrated requirements ensuring that data assets are stored, arranged, managed, and used in systems in support of organizational strategy



Trusted Catalog



Business Glossary

- aka
 - data dictionary
 - data item dictionary
 - data directory
 - data catalog
 - data repository
 - metadata repository
 - data resource dictionary
 - data asset dictionary
 - data definition dictionary
 - data structure dictionary
 - data element dictionary
 - enterprise repository
 - term bank
- Start of Enterprise Taxonomy
- Defines Initial Entities for Conceptual Data Model
- Engages the Business Community to Validate Entities and provide meaningful, agreed upon business definitions

Entity	Description	Domain Area
Donor	Funder	Business Development
Solicitations	Need for Work	Business Development
Solicitations Proposal	Response to Need for Work	Business Development
Pre-Positioning	Intelligence Gathering	Business Development
Award/Sub-Award	Funding Vehicle	Business Development
Terms Conditions	Details about a Funding Vehicle	Business Development
Budget	Amount of Money Available	Business Development
Work Plan	Set of Activities to Complete	Business Development
MPP	Monitoring Plan for Activities	Business Development
Project	An NGO Project is defined as a self-contained set of interventions or activities with the following characteristics: a) an external client; b) purchase order, contract or agreement; c) expected deliverables, outcomes and results; d) a beginning and end date of implementation; e) an approved budget; and full and/or part time NGO staff	Project Management
Geographic Area	Location in which a Central Office resides	Project Management
Office Locations		Project Management
Project Roles		Project Management
Project Artifacts		Project Management
Project Budget		Project Management
Project Work Plan		Project Management
Milestones	Schedule of completed activities	Project Management
Monitoring	Plan to measure Activities	Project Management
Evaluation	Assessment of Activities	Project Management
Indicators	Target of Outcome	Project Management
Outcomes	Statement of what needs to be accomplished	Project Management
Acct Receivable	Payments to NGO	Financial Management
Chart of Accounts	Defined Accounts	Financial Management
Payroll	Process to Pay Worker	Financial Management
Supplier	Provider of Goods or Service	Financial Management
Contract	Binding Agreement	Financial Management
Purchase Order	Statement of Good or Service	Financial Management
Performance	Level of Success	Talent Management
Benefits		Talent Management
Skills		Talent Management
Worker	Person who has been hired by NGO	Talent Management
Candidate	Potential hire of NGO	Talent Management



Trusted Catalog



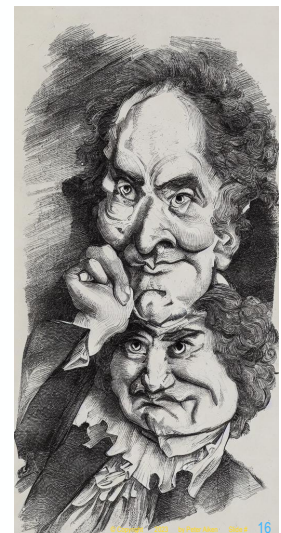
Sand in the Machinery



Program Overview

- Telling a Story (Star Wars)
- Common Vocabulary = Smooth Operations
- Objective Treatment of Governed Assets
- Data Ownership Challenges
- Data Ethics Tales
- Increasing Employee Data Literacy (Knowledge Workers)
- Productivity improvements (ChemPhDs)
- Covid undercounting
- Effective communication (Seven Sisters)
- Cultural prerequisites (Chocolate Story)

*Selling Data Investments:
Gaining Business &
Executive Support for
Data Management*



Data Assets Win!

Asset: A resource controlled by the organization as a result of past events or transactions and from which future economic benefits are expected to flow [Wikipedia]

- Today, data is the most powerful, yet underutilized and poorly managed organizational asset
- Data is your
 - Sole (only)
 - Non-depletable
 - Non-degrading
 - Durable
 - Strategic
- Asset
 - Data is the new oil!
 - Data is the new (s)oil!
 - Data is the new bacon!
- As such, data deserves:
 - It's own strategy
 - Attention on par with similar organizational assets
 - Professional ministrations to make up for past neglect

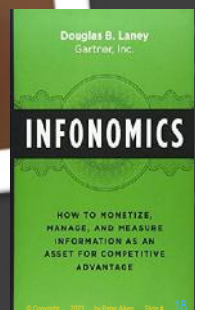
	Data Assets	Financial Assets	Real Estate Assets	Inventory Assets
Non-depletable	Available for subsequent use	Can be used up		Can be used up
Non-degrading	✓	✓	Can degrade over time	Can degrade over time
Durable	Non-taxed		✓	✓
Strategic Asset	✓	✓	✓	✓

2020 American Airlines market value ~ \$6b
 AAdvantage valued between \$19.5-\$31.5
 2020 United market value ~ 9\$b
 MileagePlus ~ \$22b

<https://www.forbes.com/sites/advisor/2020/07/15/how-airlines-make-billions-from-monetizing-frequent-flyer-programs/?sh=66da87a614e9>

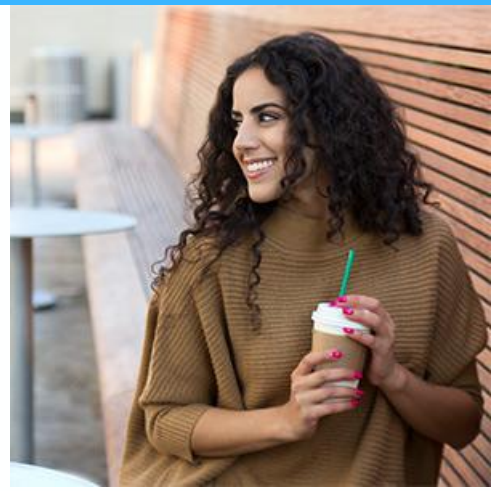


Managing Assets



Is data treated as an asset?

1. Do you have **executive positions** to support data as an asset?
2. Does the organization **track** usage of this asset?
3. Are organizational or fiscal **controls** in place to manage this asset?
4. By and large, are these controls actually **executed**?
5. Is there **general acceptance** of the need to manage this asset? (i.e., do people "get it"?)
6. Do serious discussions about this asset feature on the **agenda** of senior management meetings? [Source: Dr. Christopher Bradley]



<https://www.rooseveltdental.com/why-your-teeth-should-be-treated-as-an-asset/>



© Copyright 2023 by Peter Allen Slide 6 19

7 Data Governance Definitions

- The formal orchestration of people, process, and technology to enable an organization to leverage data as an enterprise asset – **The MDM Institute**
- A convergence of data quality, data management, business process management, and risk management surrounding the handling of data in an organization – **Wikipedia**
- A system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods – **Data Governance Institute**
- The execution and enforcement of authority over the management of data assets and the performance of data functions – **KiK Consulting**
- A quality control discipline for assessing, managing, using, improving, monitoring, maintaining, and protecting organizational information – **IBM Data Governance Council**
- Data governance is the formulation of policy to optimize, secure, and leverage information as an enterprise asset by aligning the objectives of multiple functions – **Sunil Soares**
- The exercise of authority and control over the management of data assets – **DM BoK**



© Copyright 2023 by Peter Allen Slide 6 20

Elevator Pitch



An **elevator pitch**, **elevator speech**, or **elevator statement** is a short description of an idea, product, or company that explains the concept in a way such that any listener can understand it in a short period of time.
(Wikipedia)



© 2019 Anytingawillone.com

What is Data Governance?

Managing Data with Guidance

Go Ask
Anyone!™

Would you want your sole, non-depletable, non-degrading, durable, strategic asset managed without guidance?



21

What is Data Governance?

Go Ask Anyone!™

Managing Data Decisions with Guidance

Would you want your sole, non-depletable, non-degrading, durable, strategic asset managed without guidance?



What We Learned From Top Execs About Their Big Data And AI Initiatives



Randy Bean Contributor
CIO Network Contributor Group
Enterprise & Cloud

Culture still eats strategy for breakfast

This aphorism is attributed to legendary management consultant Peter Drucker, and it certainly appears to hold true for data transformation efforts. Leading companies continue to identify culture – people, process, organization, change management – as the biggest impediment to becoming data-driven organizations – 92.2%. Few companies – only 30.0% – have even developed a well-articulated data strategy that culture could eat for breakfast. There is a lot of work to be done.

100%

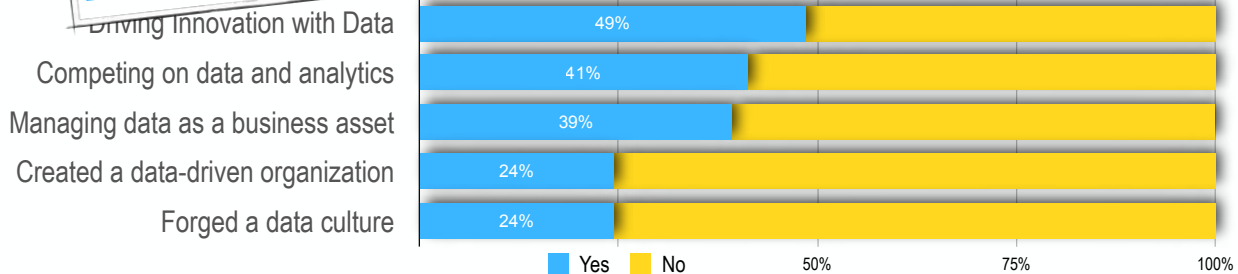
75%

92.2%

80% of data challenges are people/process based!
&
DG is the only resource to address these challenges

technology people/process

2021



Source: Big Data and AI Executive Survey 2021 by Randy Bean and Thomas Davenport www.newantage.com

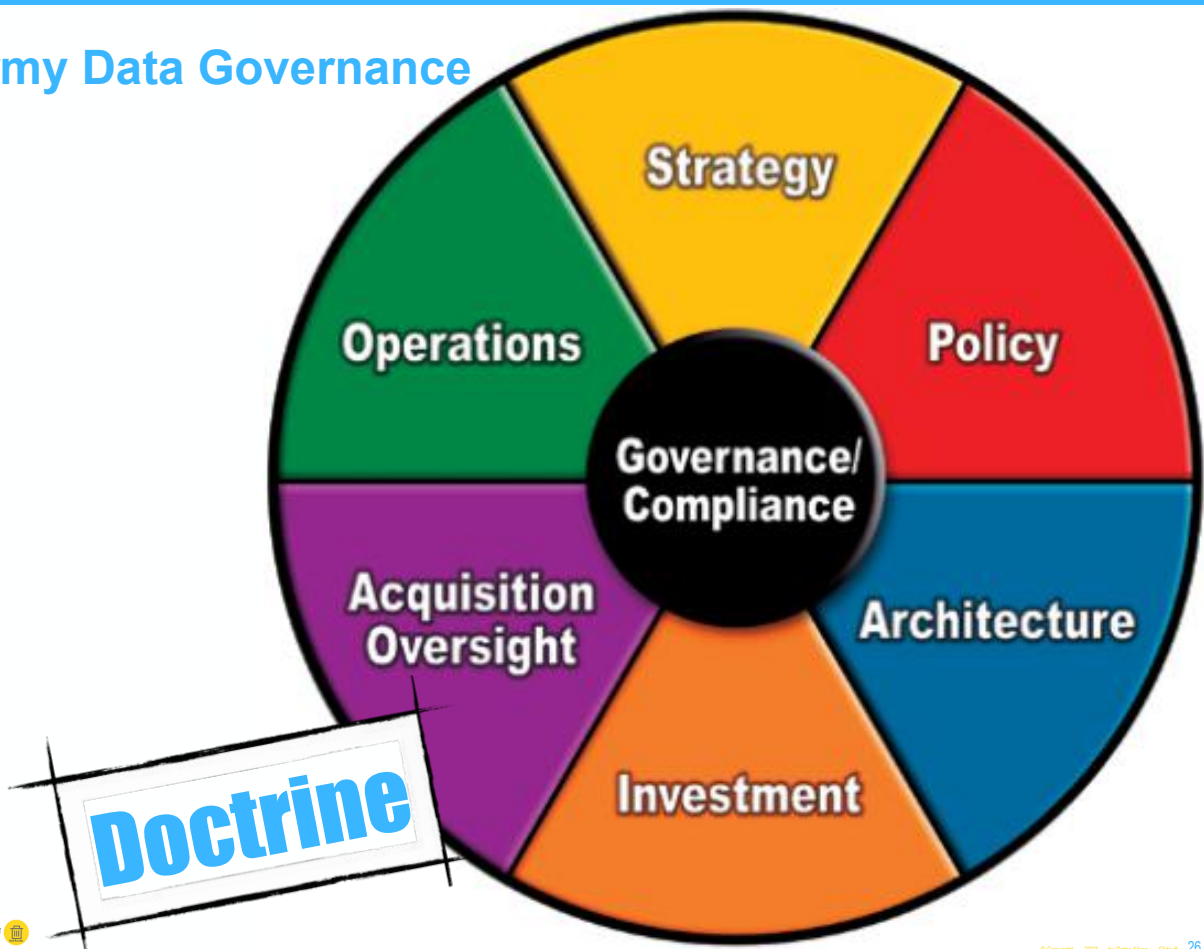
Program overview

Selling Data Investments: Gaining Business & Executive Support for Data Management

- Telling a Story (Star Wars)
- Common Vocabulary = Smooth Operations
- Objective Treatment of Governed Assets
- Data Ownership Challenges
- Data Ethics Tales
- Increasing Employee Data Literacy (Knowledge Workers)
- Productivity improvements (ChemPhDs)
- Covid undercounting
- Effective communication (Seven Sisters)
- Cultural prerequisites (Chocolate Story)



Army Data Governance



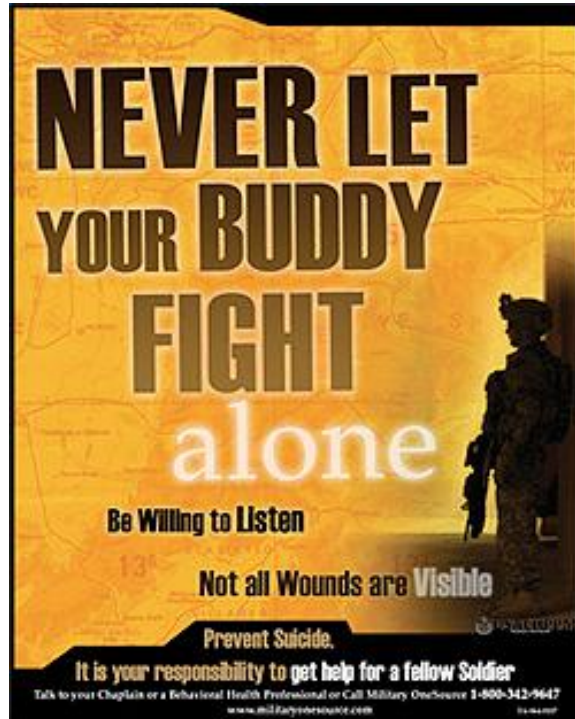
Despite prevention efforts, U.S. military suicides rise

By Halimah Abdullah | McClatchy Newspapers

WASHINGTON — Eight years of war in Afghanistan and Iraq have etched indelible scars on the psyches of many of the nation's servicemen and women, and the U.S. military is losing a battle to stem an epidemic of suicides in its ranks.

Despite calls by top Pentagon officials for a sea change in attitudes about mental health, millions of dollars in new suicide prevention programming and thousands of hours spent helping soldiers suffering from what often are euphemistically dubbed "invisible wounds," the military is losing ground.

The Department of Defense Friday reported that there were 160 reported active-duty Army suicides in 2009, up from 140 in 2008. Of these, 114 have been confirmed, while the manner of death in the remaining 46 remains to be determined.



SUICIDE AND SUICIDE-RELATED BEHAVIORS

RISK FACTORS

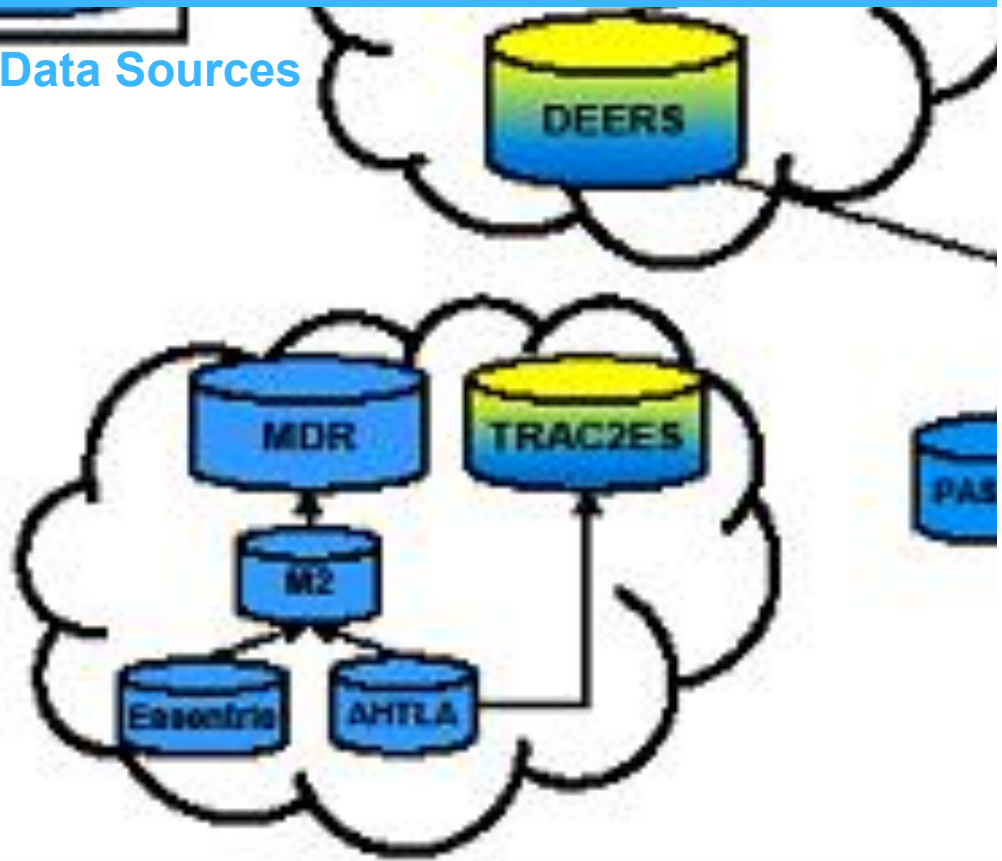
- Occupational/Command Challenges
- Financial/Legal Pressures & Stress
- Physical/Medical Illness or Pain
- Social Withdrawal & Isolation
- Marital/Relationship Issues
- Loss, Shame, Guilt & Anger
- Psychological Injury
- Emotional Distress
- Substance Abuse
- Mental Disorders
- Spiritual Conflict
- Etc.

RESILIENCY **STIGMA**

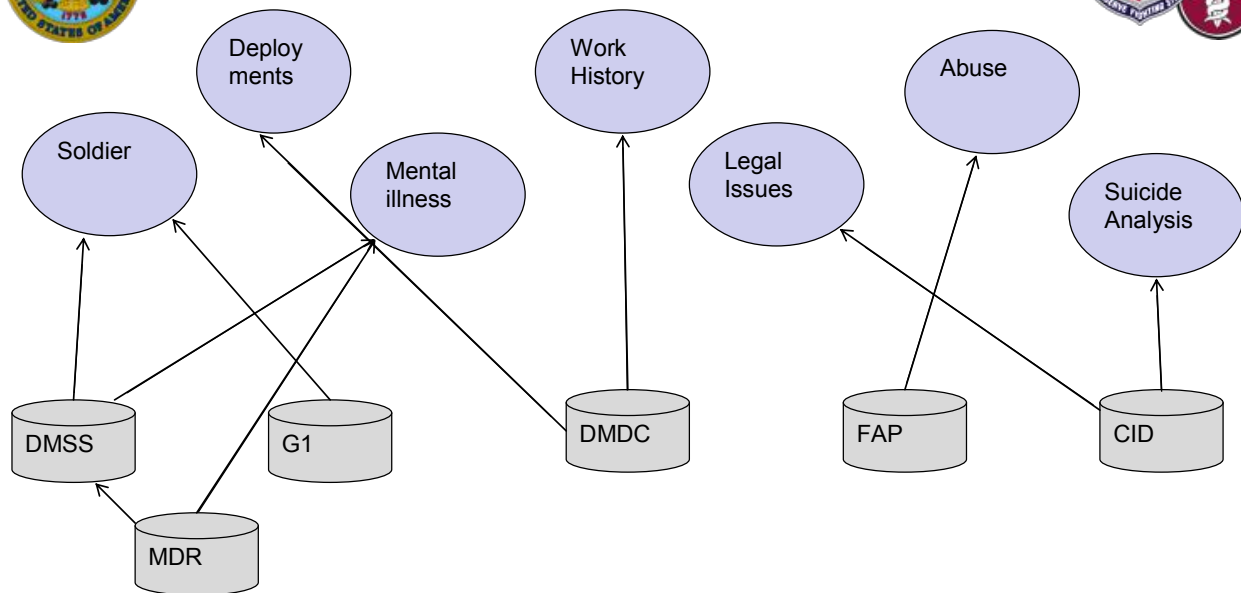
Source: The Challenge and the Promise: Strengthening the Force, Preventing Suicide and Saving Lives - The Final Report of the Department of Defense Task Force on the Prevention of Suicide by Members of the Armed Forces - August 2010



Potential Data Sources



Data Mapping



Data objects complete?

All sources identified?

Best source for each object?

How reconcile differences between sources?

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1																														
2																														
3																														
4																														
5																														
6																														
7																														
8																														
9																														
10																														
11																														
12																														
13																														
14																														
15																														
16																														
17																														
18																														
19																														
20																														
21																														
22																														
23																														
24																														
25																														
26																														
27																														
28																														
29																														
30																														



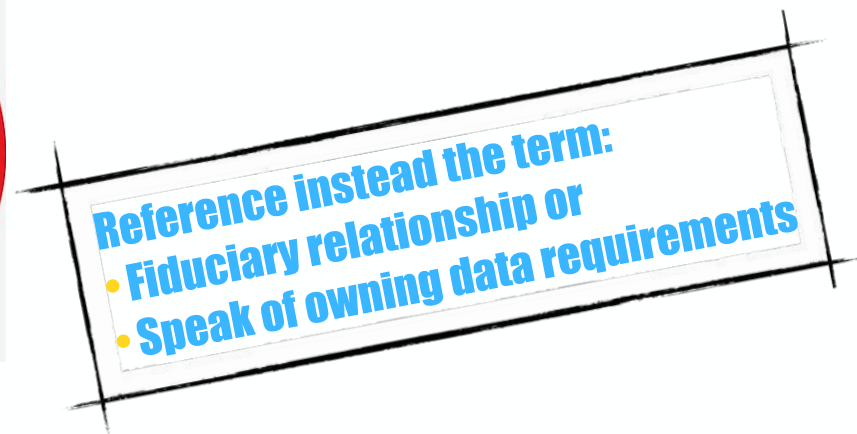
Senior Army Official

- Room full of Stewards
- A very heavy dose of management support
- Advised the group of his opinion on the matter
- Any questions as to future direction
 - "They should make an appointment to speak directly with me!"
- Empower the team
 - The conversation turned from "can this be done?" to "how are we going to accomplish this?"
 - Mistakes along the way would be tolerated
 - Implement a workable solution in prototype form



<https://anythingispossible.com>

© Copyright 2023 by Peter Allen Slide # 33



What data would accounting own?



<https://anythingispossible.com>

© Copyright 2023 by Peter Allen Slide # 34

Program overview

- Telling a Story (Star Wars)
- Common Vocabulary = Smooth Operations
- Objective Treatment of Governed Assets
- Data Ownership Challenges
- **Data Ethics Tales**
- Increasing Employee Data Literacy (Knowledge Workers)
- Productivity improvements (ChemPhDs)
- Covid undercounting
- Effective communication (Seven Sisters)
- Cultural prerequisites (Chocolate Story)

*Selling Data Investments:
Gaining Business &
Executive Support for
Data Management*



Digital/Social Reputation

"relative proximity" of three factors

- Who you **ARE**
 - Your identity in the real life, or your products and services if you run a business, the way you operate
- who **YOU SAY** you are
 - The image you wish to display to the public
- who **PEOPLE SAY** you are
 - The way others describe you or your business
- IRL blends with online?

<http://www.social-digital-mentors.eu/index.php/4-what-is-the-digital-reputation>



SOCIAL MEDIA FACTS




SOCIAL NETWORKS & BLOGS NOW ACCOUNT FOR NEARLY A QUARTER OF THE TOTAL TIME AMERICANS SPEND ONLINE. ⁴



91% OF LOCAL SEARCHES SAY THEY USE facebook TO FIND LOCAL BUSINESSES ONLINE. ⁷

RIGHT & WRONG

The Trolley Problem

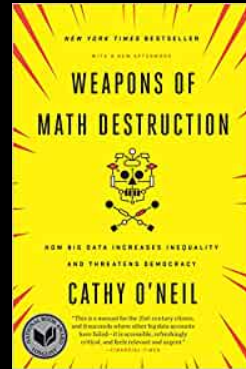
Narrated by  **Harry Shearer**



<https://www.youtube.com/watch?v=bOpf6KcWYyw> (1:30)



"Should I-thinking" must replace "Can I-thinking"
 (First level capable of being bound by defined behaviors)



Data Ethics



Data Ethics Canvas

<https://theodi.org/article/data-ethics-canvas/>

Data sources Name/describe your project's key data sources, whether you're collecting data yourself or accessing via third parties. Is any personal data involved, or data that is otherwise sensitive?	Limitations in data sources Are there limitations that could influence your project's outcomes? — bias in data collection, inclusion/exclusion, analysis, algorithms — gaps or omissions in data — provenance and data quality — other issues affecting decisions, such as team composition	Sharing data with others Are you going to be sharing data with other organisations? If so, who? Are you planning to publish any of the data? Under what conditions?	Ethical and legislative context What existing ethical codes apply to your sector or project? What legislation, policies, or other regulation shape how you use data? What requirements do they introduce? Consider: the rule of law; human rights; data protection; IP and database rights; anti-discrimination laws; and data sharing, policies, regulation and ethics codes/frameworks specific to sectors (eg health, employment, taxation).	Rights around data sources Where did you get the data from? Is it produced by an organisation or collected directly from individuals? Was the data collected for this project or for another purpose? Do you have permission to use this data, or another basis on which you're allowed to use it? What ongoing rights will the data source have?
Your reason for using data What is your primary purpose for collecting and using data in this project? What are your main use cases? What is your business model? Are you making things better for society? How and for whom? Are you replacing another product or service as a result of this project?	Communicating your purpose Do people understand your purpose — especially people who the data is about or who are impacted by its use? How have you been communicating your purpose? Has this communication been clear? How are you ensuring more vulnerable individuals or groups understand? How are you ensuring more vulnerable individuals or groups understand?	Positive effects on people Which individuals, groups, demographics or organisations will be positively affected by this project? How? How are you measuring and communicating positive impact? How could you increase it?	Negative effects on people Who could be negatively affected by this project? Could the way that data is collected, used or shared cause harm or expose individuals to risk of being re-identified? Could it be used to target, profile or prejudice people, or unfairly restrict access (eg exclusive arrangements)? How are limitations and risks communicated to people? Consider: people who the data is about, people impacted by its use and organisations using the data.	Minimising negative impact What steps can you take to minimise harm? How could you reduce any limitations in your data sources? How are you keeping personal and other sensitive information secure? How are you measuring, reporting and acting on potential negative impacts of your project? What benefits will these actions bring to your project?
Engaging with people How can people engage with you about the project? How can people correct information, appeal or request changes to the product/service? To what extent? Are appeal mechanisms reasonable and well understood?	Openness and transparency How open can you be about this project? Could you publish your methodology, metadata, datasets, code or impact measurements? Can you ask peers for feedback on the project? How will you communicate it internally? Will you publish your actions and answers to this canvas openly?	Ongoing implementation Are you routinely building in thoughts, ideas and considerations of people affected in your project? How? What information or training might be needed to help people understand data issues? Are systems, processes and resources available for responding to data issues that arise in the long-term?	Reviews and iterations How will ongoing data ethics issues be measured, monitored, discussed and acted on? How often will your responses to this canvas be reviewed or updated? When?	Your actions What actions will you take before moving forward with this project? Which should take priority? Who will be responsible for these actions, and who must be involved? Will you openly publish your actions and answers to this canvas?



© Copyright 2023 by Peter Allen Slide 6 39

Topic: Ransomware

MOTHERBOARD
TECH BY VICE

Hackers Apologize to Arab Royal Families for Leaking Their Data

<https://www.vice.com/en/article/n7nw8m/conti-ransomware-hackers-apologize-to-arab-royal-families-for-leaking-their-data>

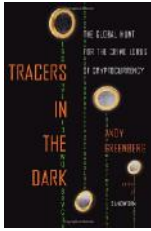
- Would you advise a criminal organization to adopt data governance?
- What advice would you give them?
- What advice wouldn't you give them?

- October 2021
 - Conti (infamous ransomware gang) released thousands of files stolen from the UK jewelry store
- Now, the hackers would like the world to know that they regret their decision, perhaps in part because they released files belonging to very powerful people. ...
 - "We found that our sample data was not properly reviewed before being uploaded to the blog," the hackers wrote in an announcement published on Thursday. "Conti guarantees that any information pertaining to members of Saudi Arabia, UAE, and Qatar families will be deleted without any exposure and review."
 - "Our Team apologizes to His Royal Highness Prince Mohammed bin Salman and any other members of the Royal Families whose names were mentioned in the publication for any inconvenience," the hackers added.
- Imagine being a big-time ransomware hacker, thinking that you're pretty tough, fancying yourself a master criminal, giving yourself an intimidating online alias, maybe even being able, in certain circumstances, to call down violence on your enemies, and then realizing one day that you'd accidentally hacked a guy who had a journalist kidnapped, tortured to death and then dismembered with a bone saw for criticizing him.
- They are adding new compliance procedures to make sure this won't happen again:
 - The hackers also said that other than publishing the data on their site, they did not sell it or trade, and that from now on they will "implement a more rigid data review process for any future operations."
- We have talked before about the compliance function at ransomware firms. If you run a legal company, you have a compliance department to make sure that you don't do anything illegal, or at least, if your company is really big, to keep the illegality within acceptable limits. If you run a criminal gang, you have concerns that are different in degree but directionally similar: Your whole business is doing illegal things, sure, but you don't want to do too many things that are too illegal. You want to do crimes that make you money, but not crimes that get you shut down. You want to steal information from rich people and extort money from them. But not Mohammed bin Salman! Good lord!



© Copyright 2023 by Peter Allen Slide 6 40

Topic: Crypto



- “Every Bitcoin user has access to the public Bitcoin blockchain and can see every Bitcoin address and its respective transfers. Due to this publicity, it is possible to determine the identities of Bitcoin address owners by analyzing the blockchain,” the ruling read. “There is no intrusion into a constitutionally protected area because there is no constitutional privacy interest in the information on the blockchain.”
- The HSI agent wasn’t caught in the Welcome to Video dragnet because IRS agents had violated his privacy. He was caught, the judges concluded, because he had mistakenly believed his Bitcoin transactions to have ever been private in the first place.



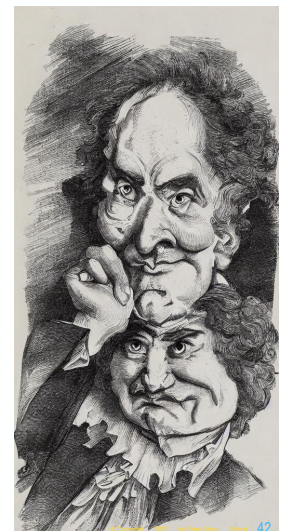
<https://www.wired.com/story/tracers-in-the-dark-welcome-to-video-crypto-anonymity-myth/>

© Copyright 2023 by Peter Allen Slide # 41

Program Overview

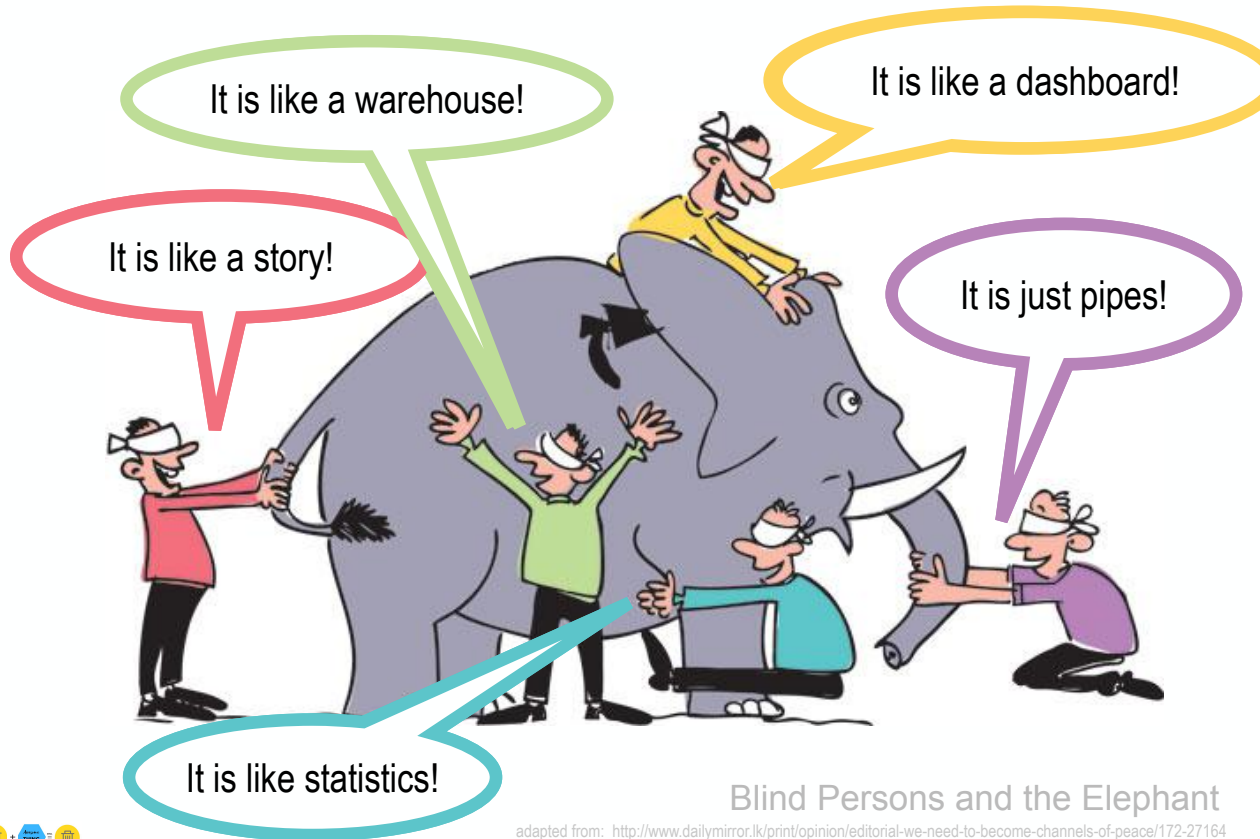
- Telling a Story (Star Wars)
- Common Vocabulary = Smooth Operations
- Objective Treatment of Governed Assets
- Data Ownership Challenges
- Data Ethics Tales
- Increasing Employee Data Literacy (Knowledge Workers)
- Productivity improvements (ChemPhDs)
- Covid undercounting
- Effective communication (Seven Sisters)
- Cultural prerequisites (Chocolate Story)

*Selling Data Investments:
Gaining Business &
Executive Support for
Data Management*



© Copyright 2023 by Peter Allen Slide # 42

Data is not broadly or widely understood



© Copyright 2023 by Peter Allen Slide 43

20 Seconds of Self-driving Software

1. Horse and buggy
2. Pedestrian
3. Semi-truck
4. Semi-truck sideways
5. Pickup truck
6. Semi-truck
7. Oncoming semi
8. Pedestrian following pickup truck
9. Pedestrian following semi-truck

Other



<https://www.dailymail.co.uk/sciencetech/article-11123757/Teslas-self-driving-software-confuses-horse-drawn-carriage-highway-semi-truck.html>

44

WHY IS IT IMPORTANT? IN A DESIGN PROCESS, WE MAKE DECISIONS BASED ON DATA!

01000100
01000001
01010100
01000001

QUANTITATIVE RESEARCH PART 2

DATA LITERACY

FOR UX DESIGNERS

00100000

UX Knowledge Base Sketch #57

01001100
01001001
01010100
01000101
01010010
01000001
01000011
01010011

CAN YOU GUESS WHAT THESE BINARY NUMBERS MEAN? :)



THE ABILITY TO CONSUME FOR KNOWLEDGE, PRODUCE COHERENTLY AND THINK CRITICALLY ABOUT DATA.

THINGS TO CONSIDER WHEN YOU WORK WITH DATA:



CONTEXT? E.G. SEASONALITY



ANY TECHNICAL ISSUES? } COLLABORATE /W DEVELOPERS! E.G. PERFORMANCE



HOW WAS IT COLLECTED / PROCESSED / VALIDATED / EDITED



SOURCE: ANY LIMITATIONS? DELAYS? RATE?



SAMPLE SIZE?



METHODOLOGY?



ANY OTHER SIGNIFICANT EXPERIMENT CONDITION? LIMITATIONS?



BE AWARE OF BIASES! E.G.:

- SELECTION BIAS
- CONFIRMATION BIAS
- SURVIVORSHIP BIAS
- HINDSIGHT BIAS
- CURSE OF KNOWLEDGE
- CLUSTERING ILLUSION AND SO ON.



CONDUCT (& DESIGN) YOUR OWN RESEARCH, COLLABORATE /W DATA SCIENTISTS IN YOUR TEAM! APPLY THE COMBINATION OF QUANTITATIVE & QUALITATIVE METHODS TO AVOID YOUR LOCAL MAXIMUM!

https://myinguide.com



NUMERACY: THE ABILITY TO UNDERSTAND & WORK WITH NUMBERS, MATHEMATICAL REASONING, INTERPRETING DATA IN NUMERICAL FORM. E.G. AVERAGES, PERCENTAGES



STATISTICAL LITERACY: TERMINOLOGY, UNDERSTANDING BASIC CONCEPTS BEING ABLE TO INTERPRET. SOME IMPORTANT CONCEPTS:



BASELINE: SERVES AS A BASIS FOR THE QUANTITATIVE RESEARCH, AN INITIAL DATASET



P < .05 STATISTICAL SIGNIFICANCE: TO PUT IT SIMPLY: "A RESULT THAT IS NOT ATTRIBUTED TO CHANCE" (JEFF SAURO)



PROBABILITY VALUE (P-VALUE): PROBABILITY OF OBSERVING DIFFERENCE IF THERE IS NO DIFFERENCE.



RANDOM (NOISE)



SAMPLING ERROR: OBSERVING A SAMPLE INSTEAD OF THE WHOLE POPULATION → INCREASING THE SAMPLE SIZE REDUCES THE CHANCE THAT WE OBSERVE RANDOM NOISE / COINCIDENCE.



SAMPLING BIAS: INCREASING THE SAMPLE SIZE REDUCES THE CHANCE THAT WE OBSERVE RANDOM NOISE / COINCIDENCE.



SOME MORE THINGS IF YOU WANT TO DIVE DEEPER (YOU CAN USE THIS SECTION AS A CHECKLIST!)



MEASURES OF CENTRAL LOCATION (MEAN, MODE, MEDIAN)



MEASURES OF DISPERSION (STANDARD DEVIATION, RANGE)



CONFIDENCE INTERVAL & MARGIN OF ERROR



ONE-TAILED & TWO-TAILED TESTS ... AND SO ON.

AND YOU'LL LEARN ABOUT MANY MORE CONCEPTS ALONG THE WAY!



DATA QUALITY: RELIABLE, COMPLETE, VALID ... DATA → USEFUL FOR THE INTENDED PURPOSES



RELATED: DATA CLEANING: IDENTIFY & CORRECT / REMOVE: DUPLICATE, INACCURATE, OUTDATED, INCOMPLETE, CORRUPT, IRRELEVANT PARTS



UNDERSTANDING THE RELATIONSHIP BETWEEN VARIABLES

CORRELATION ≠ CAUSALITY

SOMETHING IS BECAUSE OF SOMETHING ELSE

IT CAN BE A MATTER OF CHANCE / IMPACT OF A 3RD VARIABLE

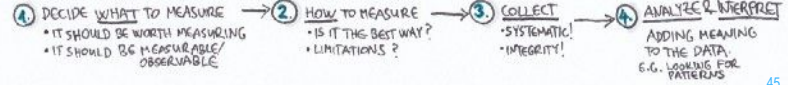


DATA TYPES: QUANTITATIVE (NUMERIC) DATA CAN BE:

- DISCRETE (E.G. INTEGER - 1, 2, 3, 4, 5)
- CONTINUOUS (ANY VALUE)

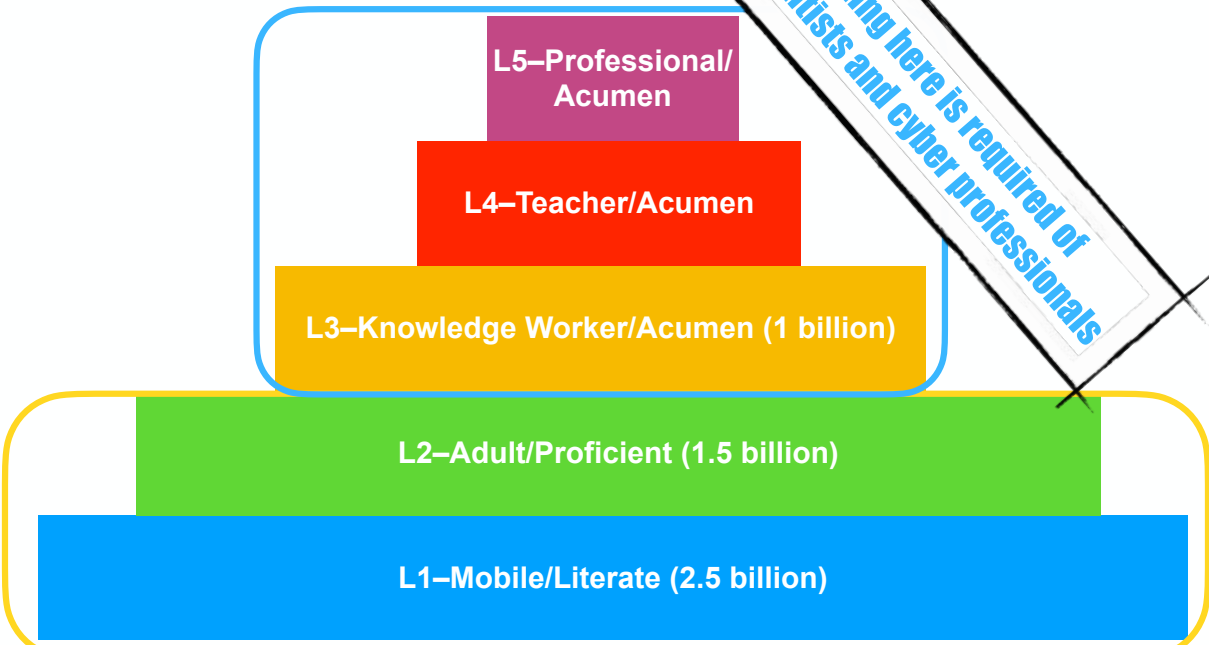
THEN MAKE DECISIONS!

PROCESS OF DATA ANALYSIS:



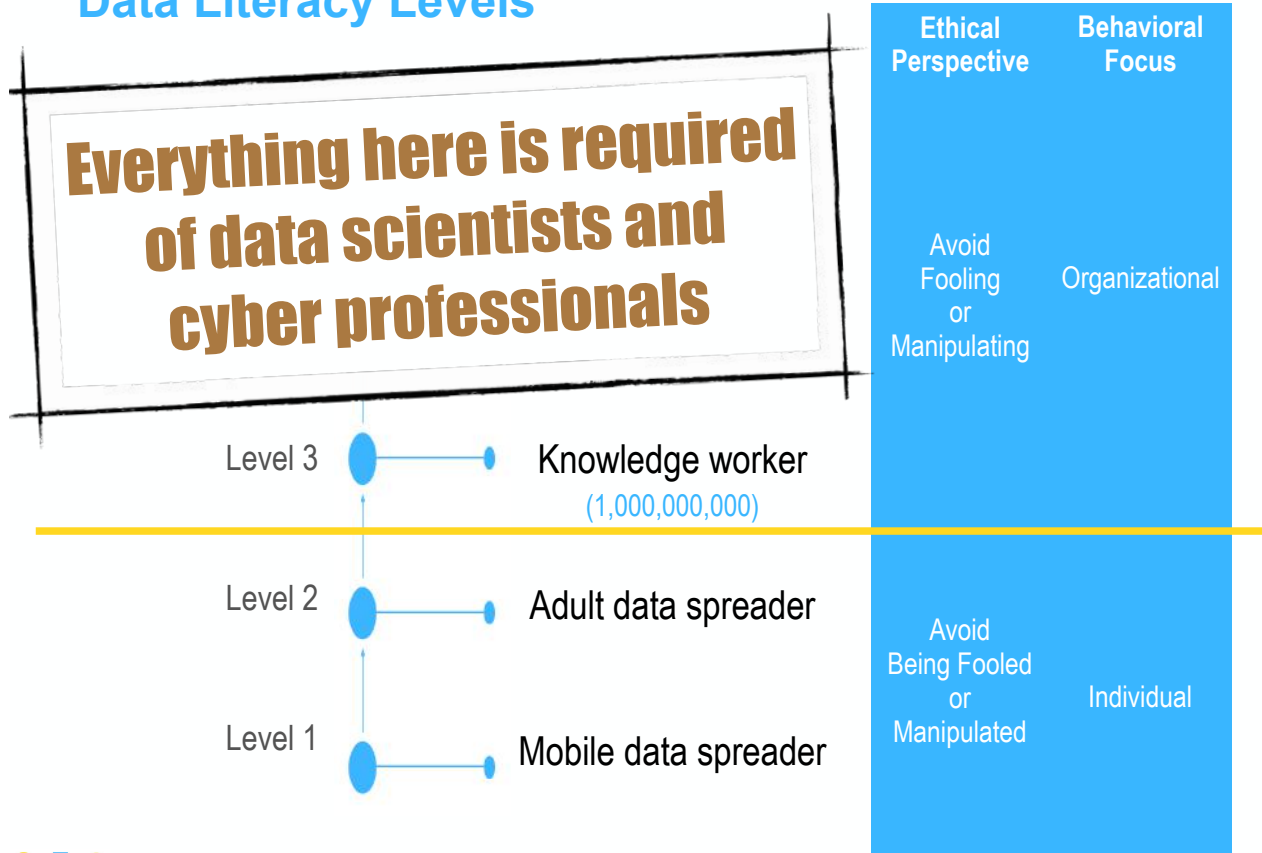
45

Digital Civics Framework



https://myinguide.com

Data Literacy Levels



A Foundation for AI Data



Whether digitizing or modernizing, garbage in–garbage out is constant. It seems such an easy concept. Yet, repeatedly we discover concerning aspects of production systems. Poor results include:

- Presenting with COVID and ASTHMA at an emergency department and receiving an evaluation of no-big-deal.
- Facial recognition systems that cannot ‘see’ certain individuals.
- Sentencing algorithms with obvious discriminatory biases in production throughout the judicial system.
- Self-driving car systems that cannot tell the difference between a semi-truck and horse-drawn carriage



Obviously, no hope exists for stamping these out individually. Instead, development practices must be revamped and roles re-conceptualized. This includes a healthy dose of data literacy and addressing the challenges remaining to fully realize potentials in this area.

- As much attention must be devoted to reverse engineering as forward engineering
- Reducing the influence of various biases via checklist-based gateway development of ethical data architectures
- Adopting value or economically focused strategies for data munging into project planning

Successfully incorporating AI into society must extend far beyond attempting to identify and categorize animals in your photo library while your phone is charging. Instead a foundation for AI data must be established.



POLITICS

Economic Impact Payments

Add Topic +

IRS sent stimulus checks to more than 1 million dead people, government watchdog agency says



Michael Collins
USA TODAY

Published 10:47 a.m. ET June 25, 2020 | Updated 12:44 p.m. ET June 25, 2020



MONEYBOX
Who Cares if the IRS Sent Stimulus Checks to Dead People?

This isn't the scandal journalists are making it out to be.

By JORDAN WEISSMANN JUNE 25, 2020 • 3:17 PM

TWEET
SHARE
COMMENT

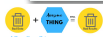


Leave the IRS alone. Zach Gibson/Getty Images

Economic Policy

Treasury sent more than 1 million coronavirus stimulus payments to dead people, congressional watchdog finds

The checks sent to dead people as of April 30 totaled nearly \$1.4 billion, according to the Government Accountability Office



Payments sent	160,000,000
to dead taxpayers	1,100,000
error rate	0.4%
(substantive?)	\$1.4 billion

Note: IRS lawyers determined "did not have legal authority to deny payments to those who filed a return for 2019, even if they were deceased at the time of payment" and checking with SSA might have slowed the distribution process down.

From SLATE:

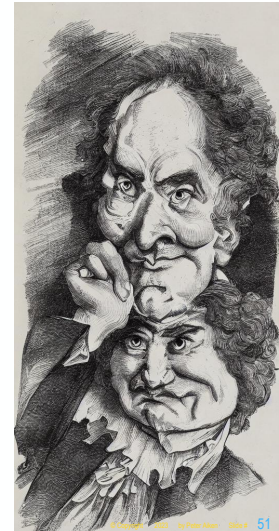
- The real headline here should really be that the government did its job pretty well!
- Speed was the priority
- Within two weeks, the IRS delivered 80 million payments electronically
- "Speed is part of the reason that poorer households were able to start spending normally again by May despite an unemployment rate rivaling the Great Depression's."



Program overview

Selling Data Investments: Gaining Business & Executive Support for Data Management

- Telling a Story (Star Wars)
- Common Vocabulary = Smooth Operations
- Objective Treatment of Governed Assets
- Data Ownership Challenges
- Data Ethics Tales
- Increasing Employee Data Literacy (Knowledge Workers)
- Productivity improvements (ChemPhDs)
- Covid undercounting
- Effective communication (Seven Sisters)
- Cultural prerequisites (Chocolate Story)



Why should a knowledge worker

- with a PhD in Chemical Engineering
- have to know whether this product was Y2K compliant?



What Is Y2K?

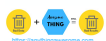
Y2K is the shorthand term for "the year 2000." Y2K was commonly used to refer to a widespread computer programming shortcut that was expected to cause extensive havoc as the year changed from 1999 to 2000. Instead of allowing four digits for the year, many computer programs only allowed two digits (e.g., 99 instead of 1999). As a result, there was immense panic that computers would be unable to operate at the turn of the millennium when the date descended from "99" to "00".

Developer's Guide



Microsoft **FOXPRO**
Relational Database Management System for MS-DOS.

I AM MORDAC, THE
PREVENTER OF INFOR-
MATION SERVICES! I
SUMMON THE Y2K DEMON!

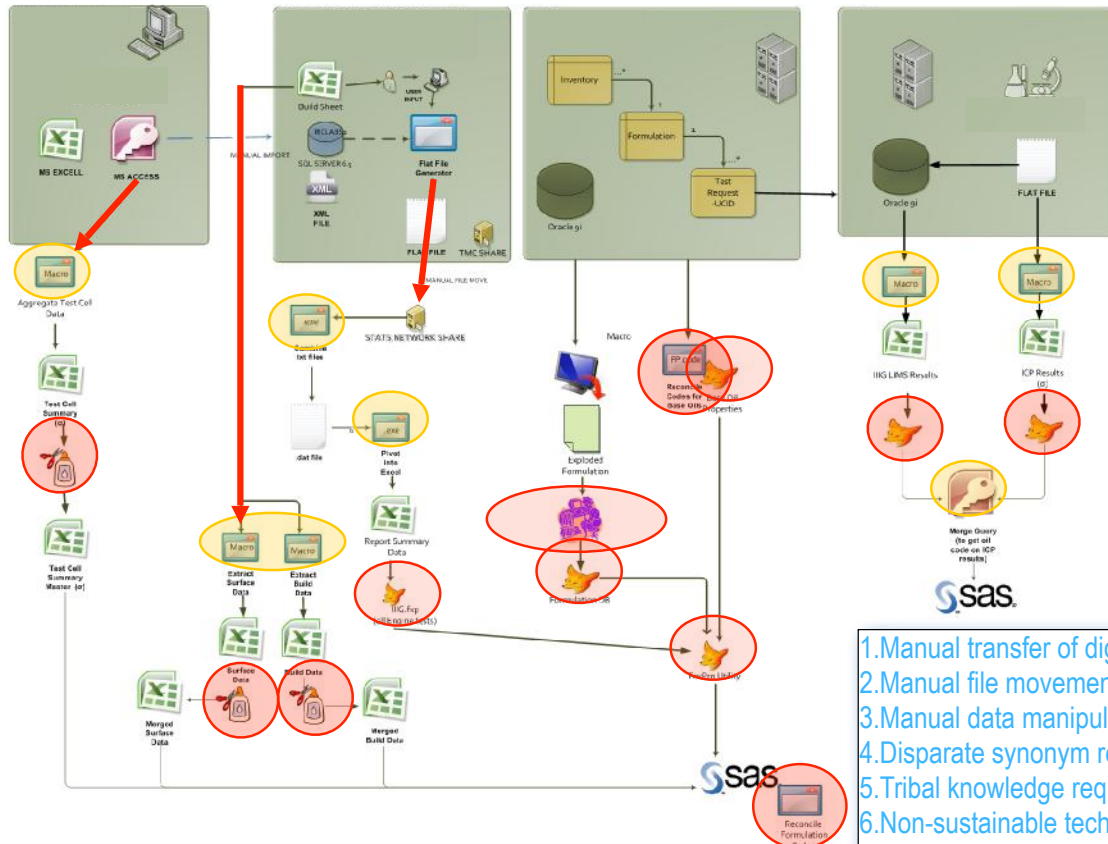


International Chemical Company Engine Testing

- \$1billion (+) chemical company
- Develops/manufactures additives enhancing the performance of oils and fuels ...
- ... to enhance engine/machine performance
 - Helps fuels burn cleaner
 - Engines run smoother
 - Machines last longer
- Tens of thousands of tests annually
 - Test costs range up to \$250,000



Improving Knowledge Worker Productivity

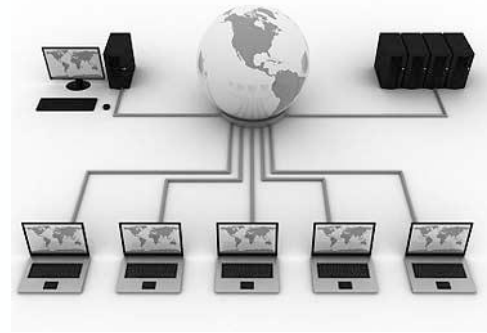


1. Manual transfer of digital data
2. Manual file movement/duplication
3. Manual data manipulation
4. Disparate synonym reconciliation
5. Tribal knowledge requirements
6. Non-sustainable technology



Data Integration Solution

- Integrated the existing systems to easily search on and find similar or identical tests
- Results:
 - Reduced expenses
 - Improved competitive edge and customer service
 - Time savings and improve operational capabilities
- According to our client's internal business case development, they expect to realize a **\$25 million gain each year** thanks to this data integration



Program Overview

- Telling a Story (Star Wars)
- Common Vocabulary = Smooth Operations
- Objective Treatment of Governed Assets
- Data Ownership Challenges
- Data Ethics Tales
- Increasing Employee Data Literacy (Knowledge Workers)
- Productivity improvements (ChemPhDs)
- Covid undercounting
- Effective communication (Seven Sisters)
- Cultural prerequisites (Chocolate Story)

*Selling Data Investments:
Gaining Business &
Executive Support for
Data Management*



Why using Microsoft's tool caused Covid-19 results to be lost



Under-reported figures

From 25 Sept to 2 Oct

50,786

Cases initially reported by PHE

15,841

Unreported cases, missed due to IT error

- Since 2007 should have been forced to use .xlsx (1,000,000+ rows)
- Used .xls (65,000 rows)
- Additional data was dropped without notification

8 days of incomplete data

1,980 cases per day, on average, were missed in that time

48 hours Ideal time limit for tracing contacts after positive test

https://www.bbc.com/news/technology-54423988?es_p=12801491

Source: PHE and gov.uk



SCIENCE | US & WORLD | TECH

Excel spreadsheet error blamed for UK's 16,000 missing coronavirus cases

The case went missing after the spreadsheet hit its filesize limit

By James Vincent | Oct 5, 2020, 9:41am EDT

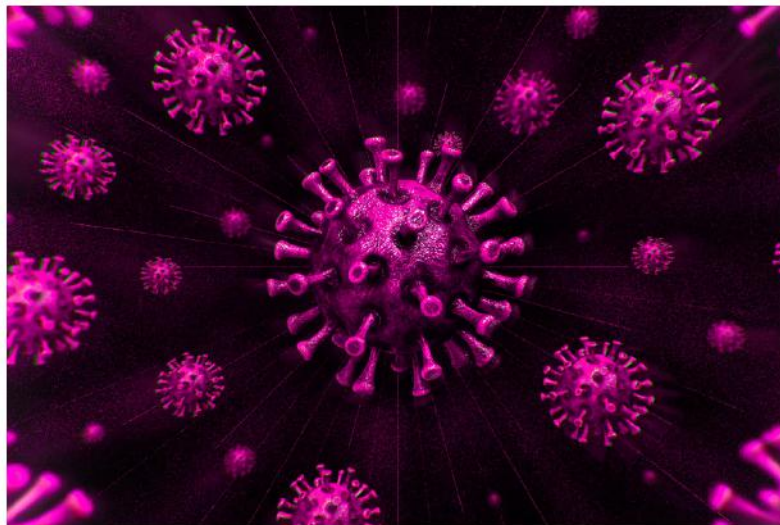


Illustration by Alex Castro / The Verge

The UK failed to add nearly 16,000 confirmed cases of coronavirus to its national track and trace system due to an Excel error. A number of reports, including from The Guardian, Sky News, and The Daily Mail, say the mistake was caused when an Excel spreadsheet used to track confirmed cases of the virus reached its maximum file size and failed to update.

"Failure to upload these cases to the national database meant anyone who came into contact with these individuals was not informed. It's an error that may have helped spread the virus further through the country as individuals exposed to the virus continued to act as normal."



Program overview

Selling Data Investments: Gaining Business & Executive Support for Data Management

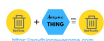
- Telling a Story (Star Wars)
- Common Vocabulary = Smooth Operations
- Objective Treatment of Governed Assets
- Data Ownership Challenges
- Data Ethics Tales
- Increasing Employee Data Literacy (Knowledge Workers)
- Productivity improvements (ChemPhDs)
- Covid undercounting
- Effective communication (Seven Sisters)
- Cultural prerequisites (Chocolate Story)



Definitions



- Planning, implementation and control activities to ensure consistency with a "golden" version of contextual data values
- ... as opposed to **mobile device management**
- Gartner holds that MDM is a **discipline** or **strategy**
 - "... where the business and the IT organization work together to ensure the uniformity, accuracy, semantic persistence, stewardship and accountability of the enterprise's official, shared master data."
- Sold as technology-based solution
- Official, consistent set of identifiers - examples of these core entities include:
 - **Parties** (customers, prospects, people, citizens, employees, vendors, suppliers, trading partners, individuals, organizations, citizens, patients, vendors, supplies, business partners, competitors, students, products, financial structures *LEI*)
 - **Places** (locations, offices, regional alignments, geographies)
 - **Things** (accounts, assets, policies, products, services)

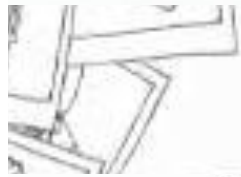




a great customer experience starts with excellent data

previously information was stored in hundreds of databases across BT

Seven Sisters (from British Telecom)



<https://anythingawesome.com/sevensistersvideo.html>

Thanks to Dave Evans

© Copyright 2023 by Peter Allen Slide 61

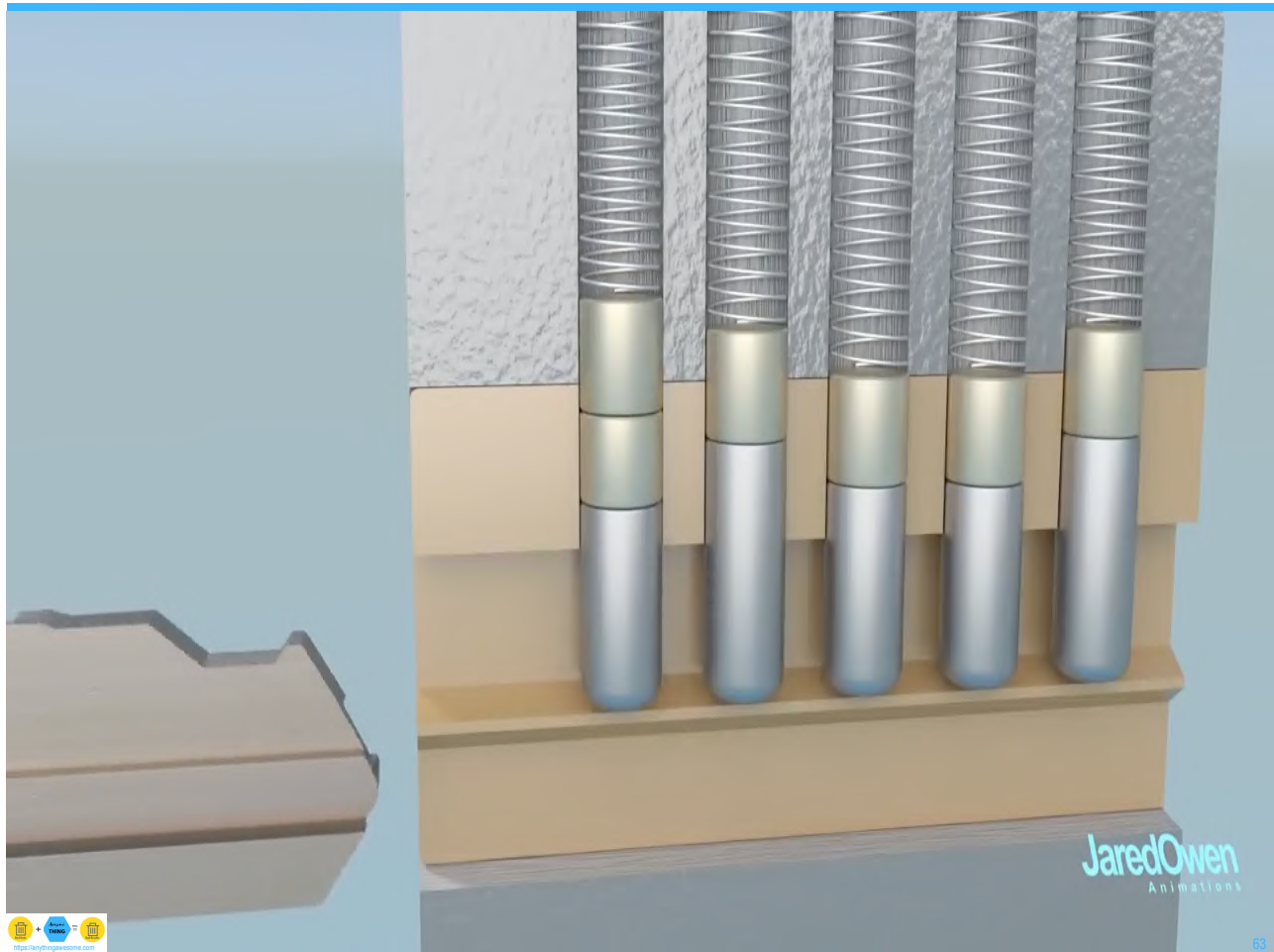
Program Overview

- Telling a Story (Star Wars)
- Common Vocabulary = Smooth Operations
- Objective Treatment of Governed Assets
- Data Ownership Challenges
- Data Ethics Tales
- Increasing Employee Data Literacy (Knowledge Workers)
- Productivity improvements (ChemPhDs)
- Covid undercounting
- Effective communication (Seven Sisters)
- Cultural prerequisites (Chocolate Story)

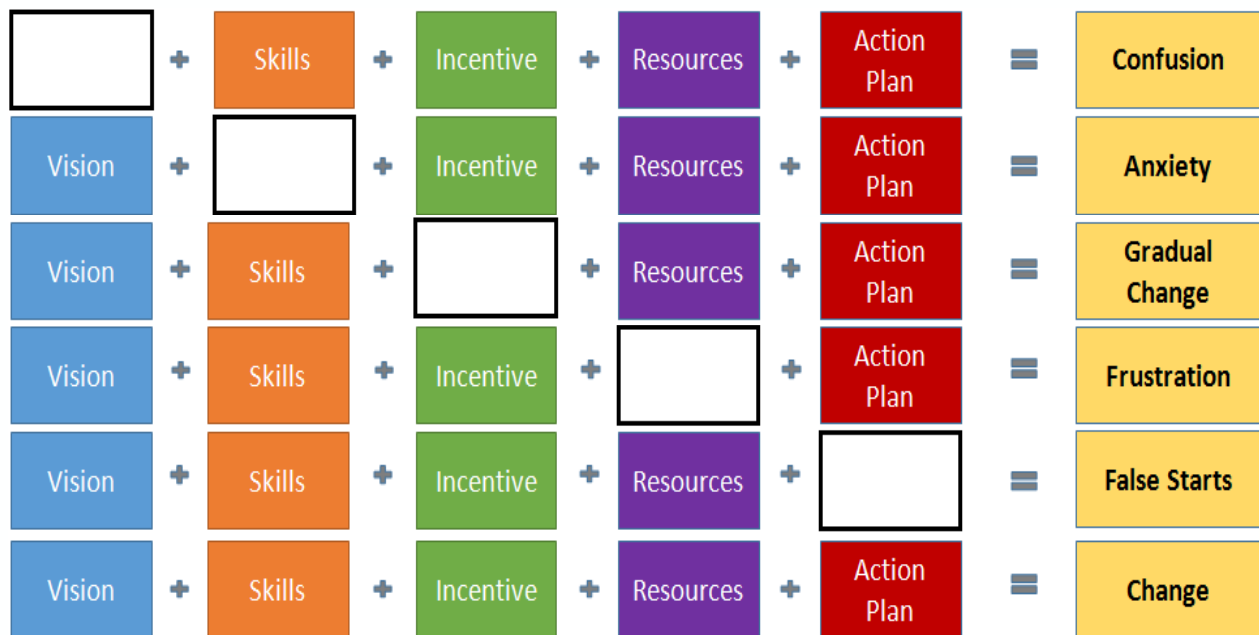
*Selling Data Investments:
Gaining Business &
Executive Support for
Data Management*



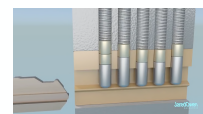
© Copyright 2023 by Peter Allen Slide 62



Diagnosing Organizational Readiness



Culture is the biggest impediment to a shift in organizational thinking about data!





[Clicking any webinar title will link directly to the registration page]

Upcoming Events

Key Elements of a Successful
Data Governance Program
13 June 2023



Conceptual vs. Logical vs.
Physical Data Modeling
11 July 2023



Time: 19:00 UTC (2:00 PM NYC) | **Presented by:** Peter Aiken, PhD

The Importance of Metadata:
3 Leveraging Strategies
8 Aug 2023

Brought to you by:



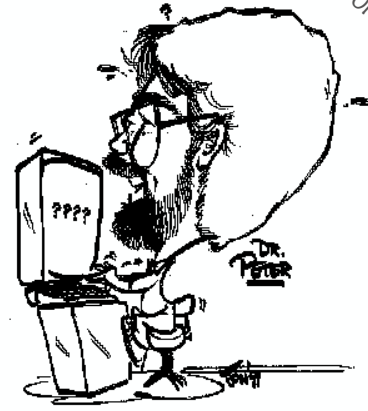
Independent Verification & Validation

Critical Design Review?

Mentoring?

Collaboration?

Executive Data Literacy Training?



Peter.Aiken@AnythingAwesome.com +1.804.382.5957

Reverse Engineering Expertise?

Hiring Assistance?

Thank You!

Use your data more strategically?

Tool/automation evaluation?



Book a call with Peter to discuss anything - <https://anythingawesome.com/OfficeHours.html>