



PROJECT MANAGEMENT  
CENTER FOR EXCELLENCE

A.J. CLARK SCHOOL OF ENGINEERING  
Civil & Environmental Engineering Department



# CLOUD BASED STRATEGIES FOR DATA MANAGEMENT AND BUSINESS INNOVATION IN PROJECTS

*Dr. Sandra Fonseca-Lind*

*2020 Project Management Symposium*



# DATA BREACH STATISTICS

DATA RECORDS LOST OR STOLEN SINCE 2013

14,717,618,286

ONLY 4% of breaches were "Secure Breaches" where encryption was used and the stolen data was rendered useless.

DATA RECORDS ARE LOST OR STOLEN AT THE FOLLOWING FREQUENCY



EVERY DAY

6,517,989



EVERY HOUR

271,583



EVERY MINUTE

4,526



EVERY SECOND

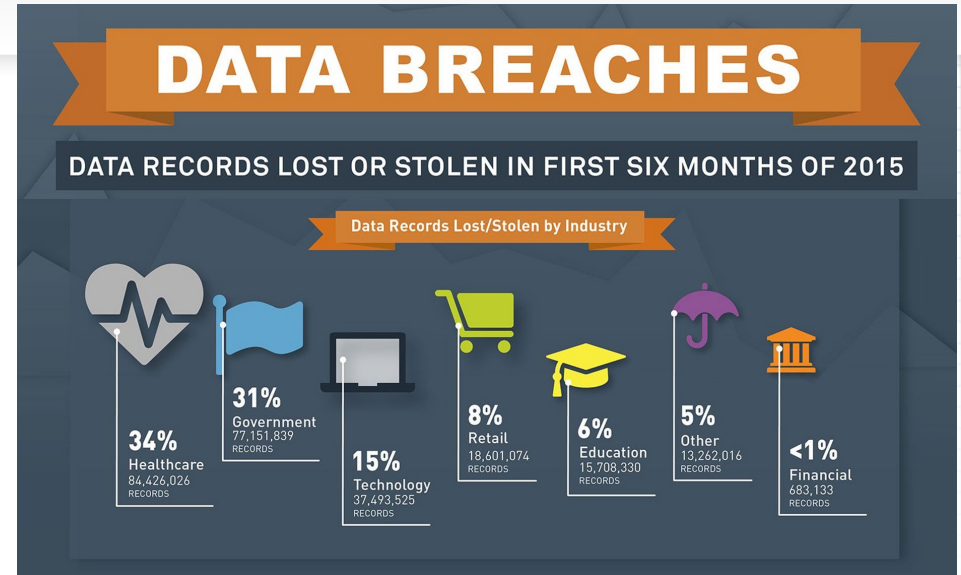
75

Source: <https://breachlevelindex.com/>

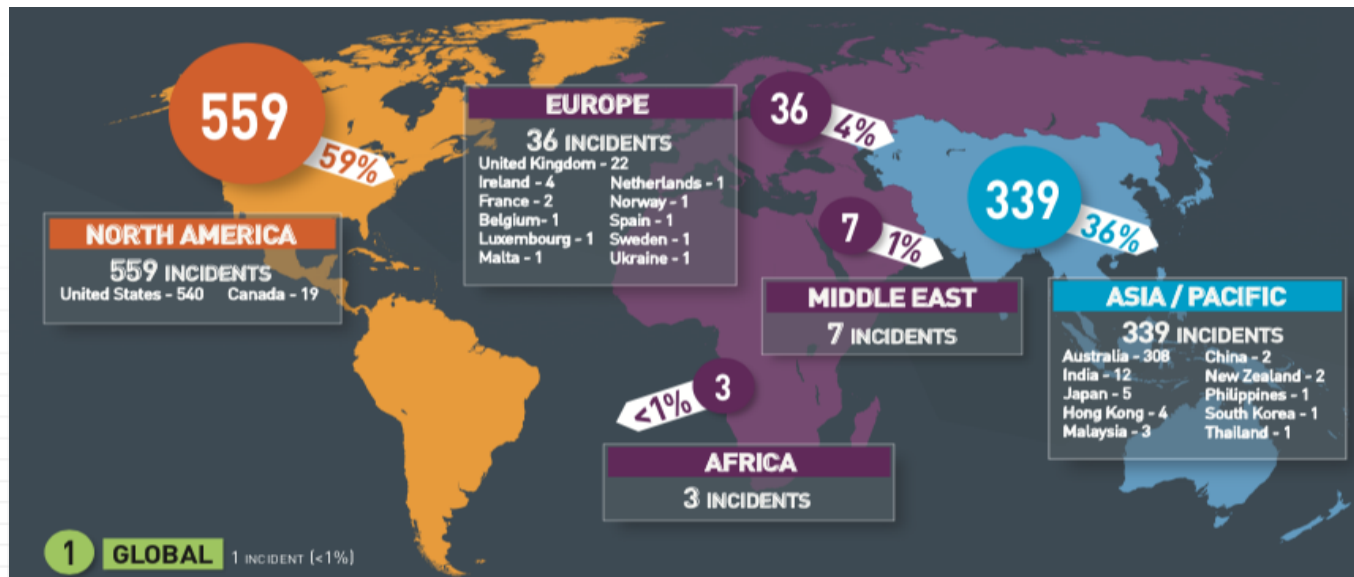
## DETECTION AND INCIDENT RESPONSE



Source: <https://smallbiztrends.com/2018/09/data-breach-statistics.html>



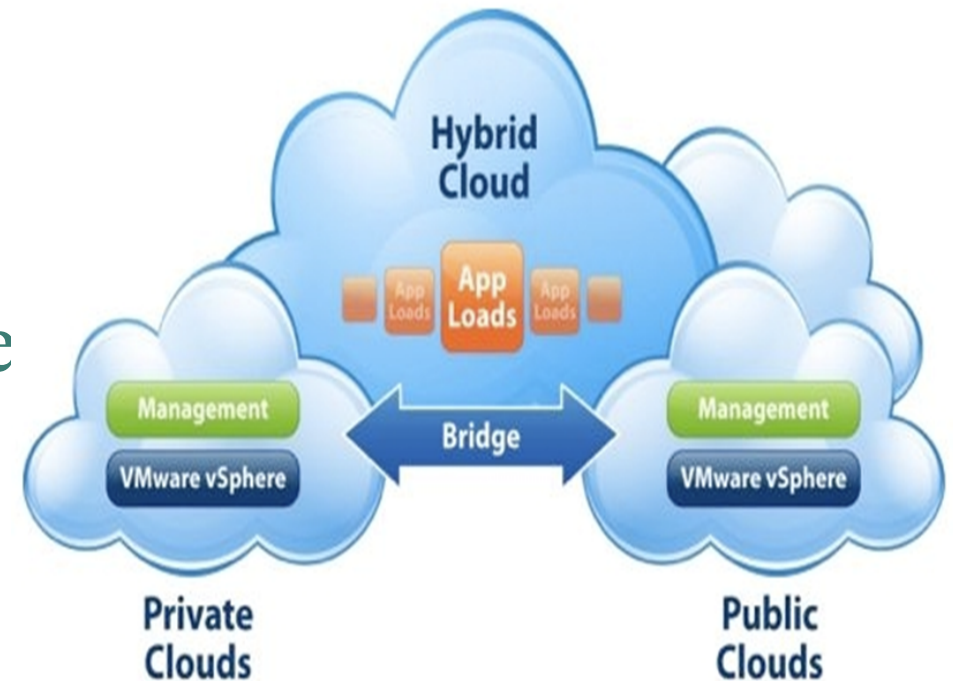
Source: <http://info.nutmegtech.com/hubfs/data-breaches-2015.jpg>



The number of records stolen in the first half of 2018 reached an astonishing **4,553,172,708**, an increase of 133 percent over the *first half of 2017* according to a report from Gemalto. (Softpedia News) - <https://news.softpedia.com/news/4-5-billion-records-stolen-in-data-breaches-in-the-first-six-months-of-2018-523174.shtml>

# Cloud Computing

- Moving computing resources out on the Internet.
- Implementing virtual servers
- Sharing data from multiple locations
- Storing and accessing applications and over the internet instead of local device
  - Reliability and Security Issues





# NIST Definition of Cloud Computing

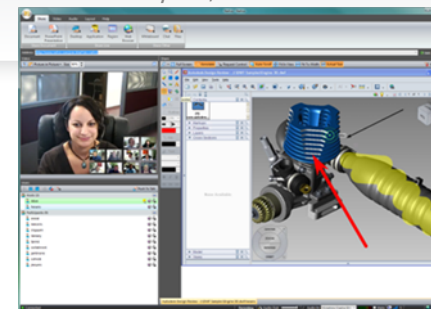
- *“Model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable resources that can be rapidly provisioned and released with minimal management effort or service provider interaction.”*

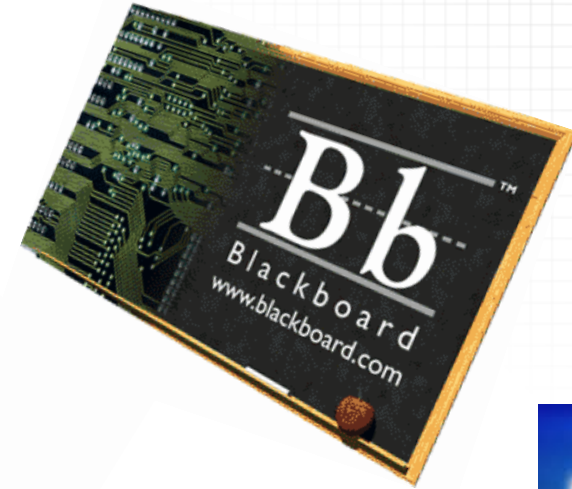
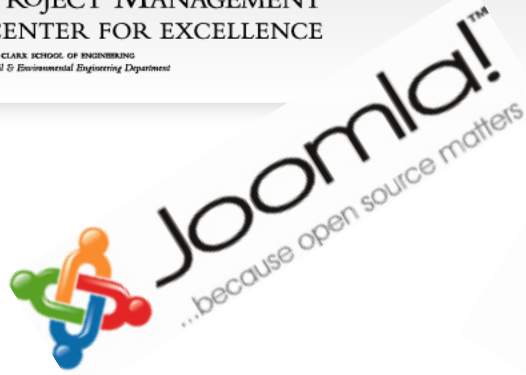


<http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf>

# Cloud Characteristics

- On-Demand Self-Service
  - Process Agility
- Broad Network Access
  - High Availability
- Resource Pooling
- Rapid Elasticity
- Measured Service





amazon  
web services



Microsoft  
Azure



canvas









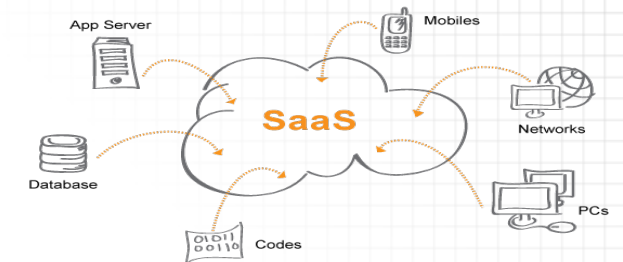
Instagram





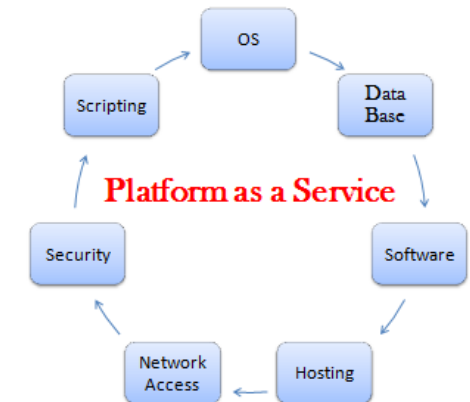
## SaaS – Software as a Service

Capability provided to the consumer to use provider's application running on a cloud infrastructure



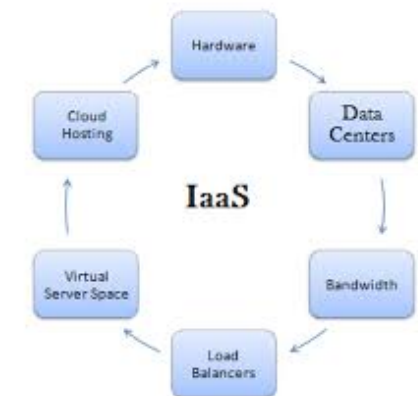
## PaaS – Platform as a Service

Capability provided to the consumer to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider.



## IaaS – Infrastructure as a Service

Capability provided to the consumer to provision processing, storage, networks, and fundamental computing resources. Consumer does not manage the infrastructure but has control over operations.

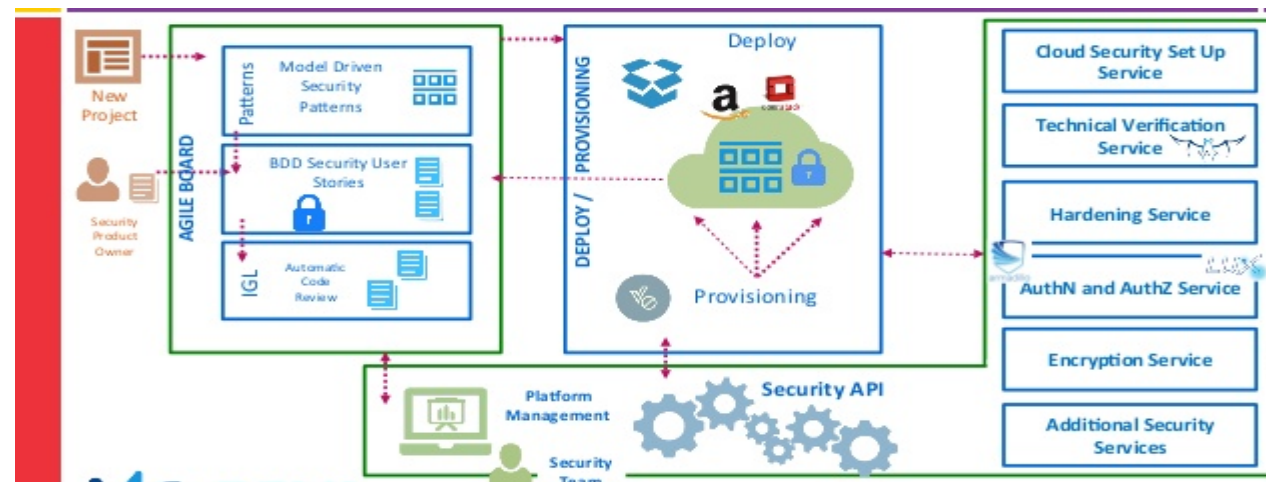


## DaaS – Data as a Service

Design, implement and manage the master data management landscape and oversee data governance compliance.

## SaaS – Security as a Service

A third party is contracted to oversee overall network perimeter and monitor for security.



# Benefits and Risks of Handling Data over the Cloud

## Benefits (Clarke, 2010)

- Availability
- Service Reliability and Integrity
- Cost Benefits associated with contracting remote resources
- Data availability at all times and in any place
- Lower costs associated with infrastructure management

## Risks (Clarke, 2010)

- Accountability
- Legal compliance
- Auditing limitations
- Compromised, lost or stolen data
- Data leaks
- System downtime from external resources
- *Project Data is part of the Enterprise History*

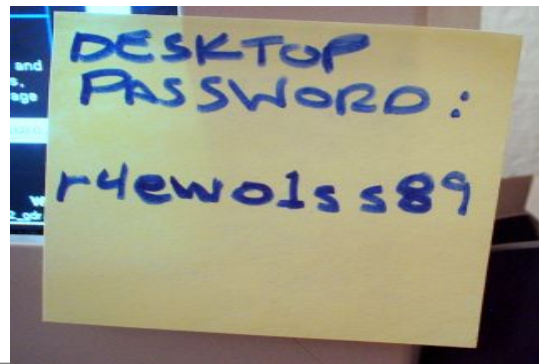
# Business and Project Management Control Strategy

- Have in place a corporate-aligned technology resources management strategies
  - Cloud Data Management
- Compliance with local and international laws and requirements
  - Accountability – Service Level Agreement (SLA)
- Protection of intellectual property – compliance evaluation
  - Service Level Agreement (SLA)
- Aligning access controls to business or institutional needs
- Sound judgment when publishing sensitive data
  - Data Encryption



# Business and Project Management Control Strategy

- Data protection before emailing
  - Passwords and Data Encryption
- Implement and manage a robust password structure
  - Two-Point Authentication
  - Three Point Authentication
- Have a backup and recovery strategy
- Use pdf-secured, passwords or digital certificates (https) when transmitting files over the cloud







# DATA GOVERNANCE FOR ENTERPRISE DATA SECURITY



# Big Data Landscape 2016

## Infrastructure

**Hadoop On-Premise**  
cloudera, Hortonworks, MAPR, Pivotal, IBM InfoSphere, splice, bluedata, jethro

**Hadoop in the Cloud**  
amazon, Microsoft Azure, Google Cloud Platform, IBM InfoSphere, CAZENA, altiscale, Du bale, xplenty

**Spark**  
databricks, GridGain, TACHYON, NEXUS

**Cluster Services**  
amazon, Kubernetes, docker, MESOSPHERE, Core OS, pepperdata, StackIQ

## Analytics

**Analyst Platforms**  
Palantir, AYASDI, Quid, enigma, Digital Reasoning, ORBITALINSIGHTS

**Analytics Platforms**  
Microsoft, guavus, Datameer, interana

**Data Science Platforms**  
context relevant, DataRobot, Alpine, MODE, plotly, ADATA, dataiku, ANTONIAN, DOMINO, sense, yhat, ALGORITHMS

**Visualization**  
tableau, Google Cloud Platform, Roambi, GOMDATA, Qlik, CHARTIO

## Applications

**Sales & Marketing**  
RADIUS, Gainsight, bloomreach, Zeta, livefyre, blueyonder, kahuna, Lattice, SAILTHRU, persado, infer, sense, AVISO, ACTIONIQ, QUANTIFIND, ZENAGGIO

**Customer Service**  
MEDALLIA, ATTENITY, CLARABRIDGE, STELLASERVICE, NGDATA, Preact, DigitalGenius, appurri, fusemachines

**Human Capital**  
gild, Connectifier, textIQ, entelo, hiQ

**Legal**  
RAVEL, JUDICATA, Everlaw, Brevia, PRESENTION

**NoSQL Databases**  
amazon DynamoDB, Google Cloud Platform, Microsoft Azure, mongoDB, MarkLogic, DATASTAX, <EROSPIKE, Couchbase, SequoiaDB, redislabs, Influxdata

**NewSQL Databases**  
SAP, Clustrix, Pivotal, paradigm4, memsql, NUODB, MariaDB, YOLDB, citusdata, deepdb, Trafalgar, Cockroach LABS

**BI Platforms**  
Power BI, amazon, Domo, Wave Analytics, GoodData, birst, GoodData, Kivus, platforma, looker, atscales, BUSINESS

**Statistical Computing**  
sas, SPSS, MATLAB

**Log Analytics**  
splunk, sumologic, kibana, CLOUD PHYSICS, loggly

**Social Analytics**  
NETBASE, DATA SIFT, tracx, bitly, synthesio, bottlenose, simple reach

**Ad Optimization**  
MediaMath, Integral, OpenX, rocketfuel, theTradeDesk, Adgorithms, Liventent, dstillery, DataXu, Appier, TARD

**Security**  
CYLANCE, CounterTack, cyberason, AREA 1 SECURITY, SenseOne, Recorded Future, FORTSCALE, sift science, Kaybase, feedzai, SCNFYD

**Vertical AI Applications**  
facebook, Clara, KASIST, lumia

**Graph Databases**  
neo4j, OrientDB, EtimeGraph

**MPP Databases**  
TERADATA, VERTICA, NETEZZA, kognitio, dremio

**Cloud EDW**  
amazon, Microsoft Azure, Pivotal, snowflake, AWS, Infoworks

**Data Transformation**  
alteryx, TRIFACTA, tamer, StreamSets, Alation

**Data Integration**  
informatica, Put potential to work, MuleSoft, snapLogic, Bedrock Data

**Real-Time**  
amazon, METAMARKETS, confluent, DATASTORY, dataArtisans

**Machine Learning**  
Amazon, H2O, SKY TREE, rapidminer, DATA GEN, DeepCense, VISENZE, PredictionIO, glowfish

**Speech & NLP**  
NarrativeScience, aiplai, NUANCE, Grindspace, semantic machines, cortico, Mindfield, IDIBON, vscor

**Horizontal AI**  
IBM Watson, Cortana, sentient, VIV, Numenta, MetaMind, clarifai

**Publisher Tools**  
outbrain, mixpanel, Chartbeat, yieldbot, Yieldmo

**Govt/ Regulation**  
Socrata, OPENGOV, EN FiscalNote, enigma, mark43, PREDPOL, OpenDataSoft

**Finance**  
affirm, LendingClub, OnDeck, Kreditech, finance, LendUp, Kabbage, tidemark, INSIKT, uora, Dataminr, Lenddo, KENSHC, AIDYA, ISENTIUM, Quantopian, sentient

**Management / Monitoring**  
New Relic, APPDYNAMICS, amazon, actifio, Numerify, splunk, DATADOG, Trocena, Anodot

**Security**  
TANIMUM, illumio, CODE42, DataGravity, CipherCloud, VECTRA, sqrl, BlueTalon

**Storage**  
amazon, Microsoft Azure, panasas, nimblestorage, Qumulo

**App Dev**  
apigee, CASK, Typesafe, CONCURRENT

**Crowd-sourcing**  
amazon, CrowdFlower, WorkFusion

**Search**  
hp, EXALTERO, ENRICA, Lucidworks, elastic, ThoughtSpot, MAANA, swifttype, Algolia, SINEQUA

**Data Services**  
UO, OPERA, Mu Sigma, DATA SCIENCE, DATA BIOMASS, kaggle, DataKind

**For Business Analysts**  
OrigamiLogic, ClearStory, CIRRO, Import IO

**SMB / Commerce**  
Google Analytics, AMPURUS, RJMetrics, BLUECORE, sumAll, granify, retention, custora

**Education/ Learning**  
KNEWTON, Clever, Geclara, PANORAMA, knowto

**Life Sciences**  
23andMe, Counsyl, Recombine, XTRUS, FLATIRON, oerazymergen, HealthTop, METABIOTA, ZEPHYRUS, ovia, Gingerio, transcriptic, Glow, enihic, AICure, Atomwise

**Industries**  
eHarmony, OPOWER, RetailNext, duetto, STITCH FIX, WorkFusion, BLUE RIVER, TACHYUS, Seeq, FarmLogs, SwiftKey, HowGood, select, statmuse, BOXEVER

## Cross-Infrastructure/Analytics

amazon, Google, Microsoft, IBM, SAP, SAS, hp, vmware, talend, TIBCO, TERADATA, ORACLE, NetApp

**Framework**  
Hadoop, YARN, Spark, MESOS, TEZ, Flink, CDAP

**Query / Data Flow**  
SLAMDATA, DRILL, Google Cloud Dataflow

**Data Access**  
cassandra, riak, mongoDB, kafka, SciDB, nifi, CouchDB, OPEN YIELD

**Coordination**  
talend, Apache Zookeeper, Apache Ambari

**Real-Time**  
STORM, Spark, Flink, druid, TAQYON

**Stat Tools**  
R, Scala, SciPy

**Machine Learning**  
mllib, Aerosolve, Apache, SINGA, Caffe, FeatureFu, DIMSUM, DL4J, VELES, jupyter

**Search**  
elasticsearch, Solr

**Security**  
Apache Ranger

**Visualization**  
Tableau

## Open Source

## Data Sources & APIs

**Health**  
Apple, JAWBONE, GARMIN, practicefusion, fitbit, Withings, VALIDIC, nefatmo, kinsa, Human API

**IOT**  
UPTAKE, ThingWorx, helium, samsara

**Financial & Economic Data**  
Bloomberg, DOW JONES, YODLEE, PREMISE, S&P CAPITAL IQ, quandl, xignite, CBINSIGHTS, mattermark, Gestimize, FLAID

**Air / Space / Sea**  
PLANET Labs, spire, WINDWARD, CRUISE, Airware, DroneDeploy, SKY CATCH

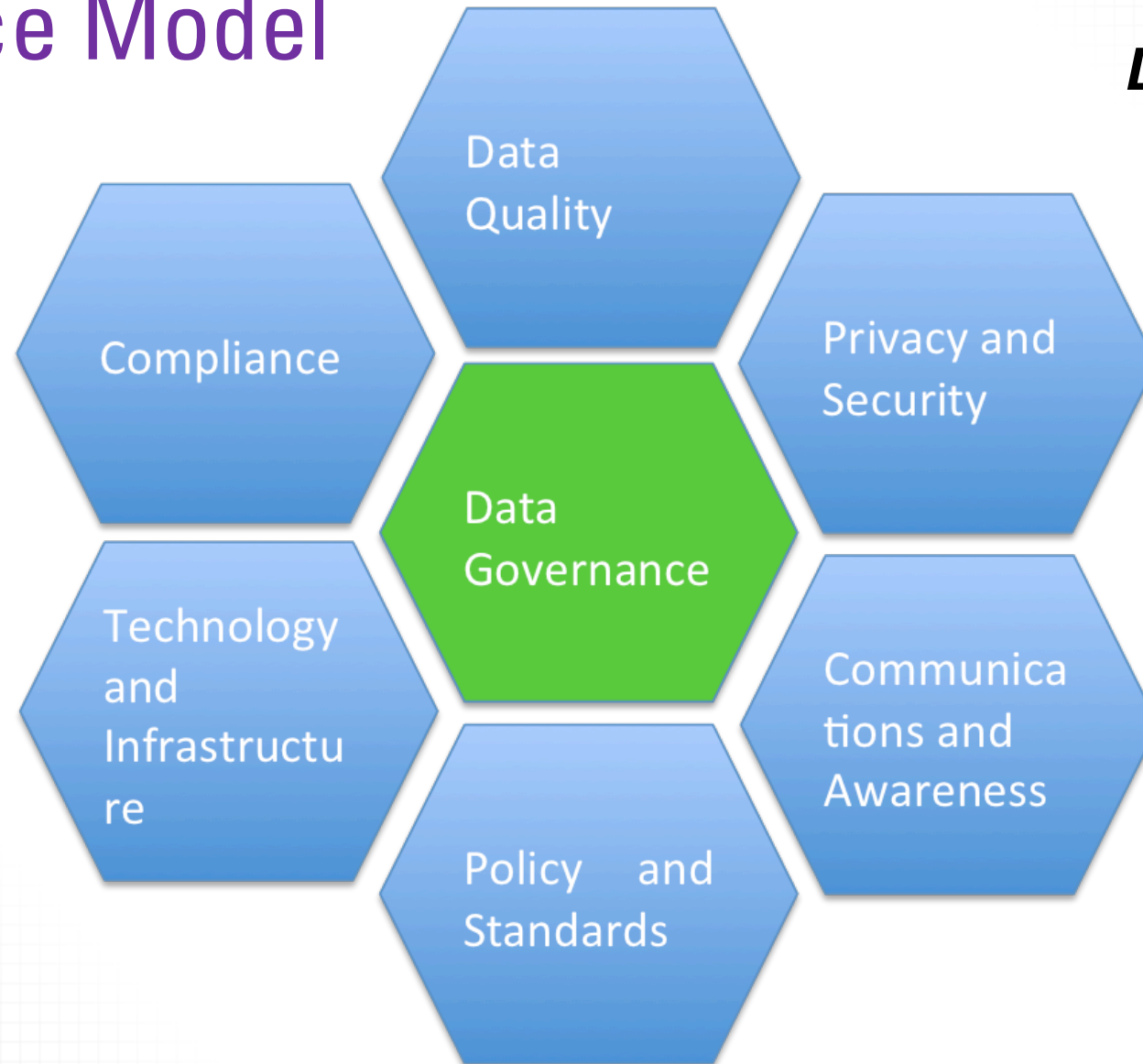
**Location/People/Entities**  
GARMIN, foursquare, InsideView, esri, STREETLINE, CARTOON, factual, PlaceIQ, Crimson Hexagon, placemeter, BASIS, Sense

**Other**  
qualtrics, panjiva, DATA GOV

**Incubators & Schools**  
GA, DataCamp, INSIGHT, DataElite, METIS, The Data Incubator



# Data Governance Model

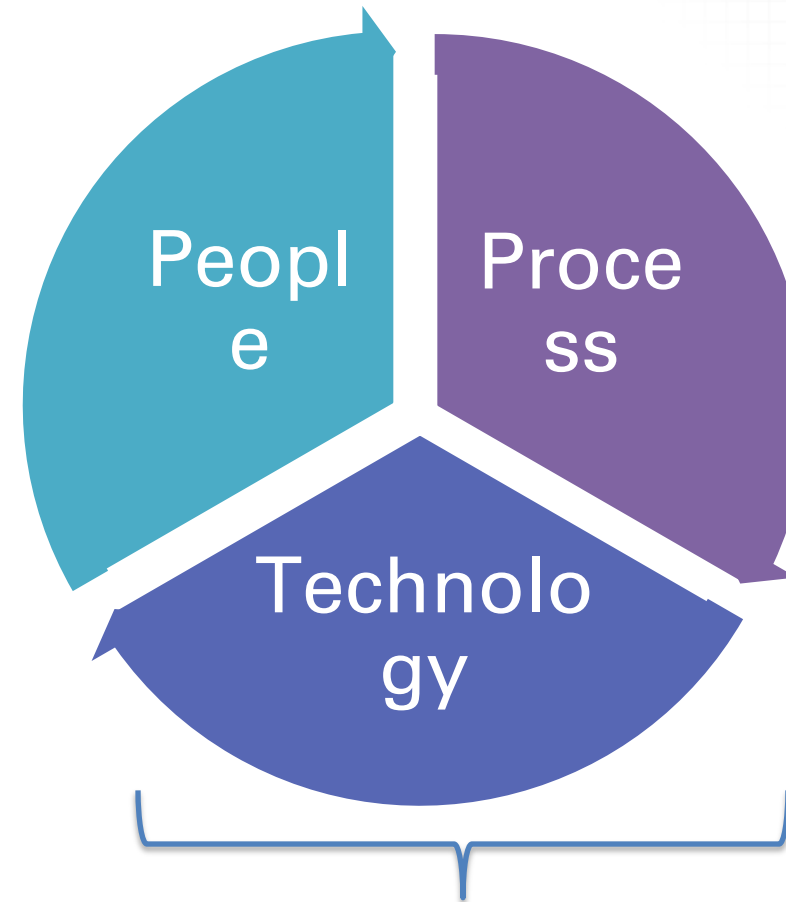


***Data Matters!***

**Data is Power!**

## Business Case

- Big Data installations with structured (DBMS) and unstructured (emails, multimedia) data stored locally and remotely via cloud services presents management challenges to manage a data centric control structure



*Data (Structured, Unstructured,  
Communications)*

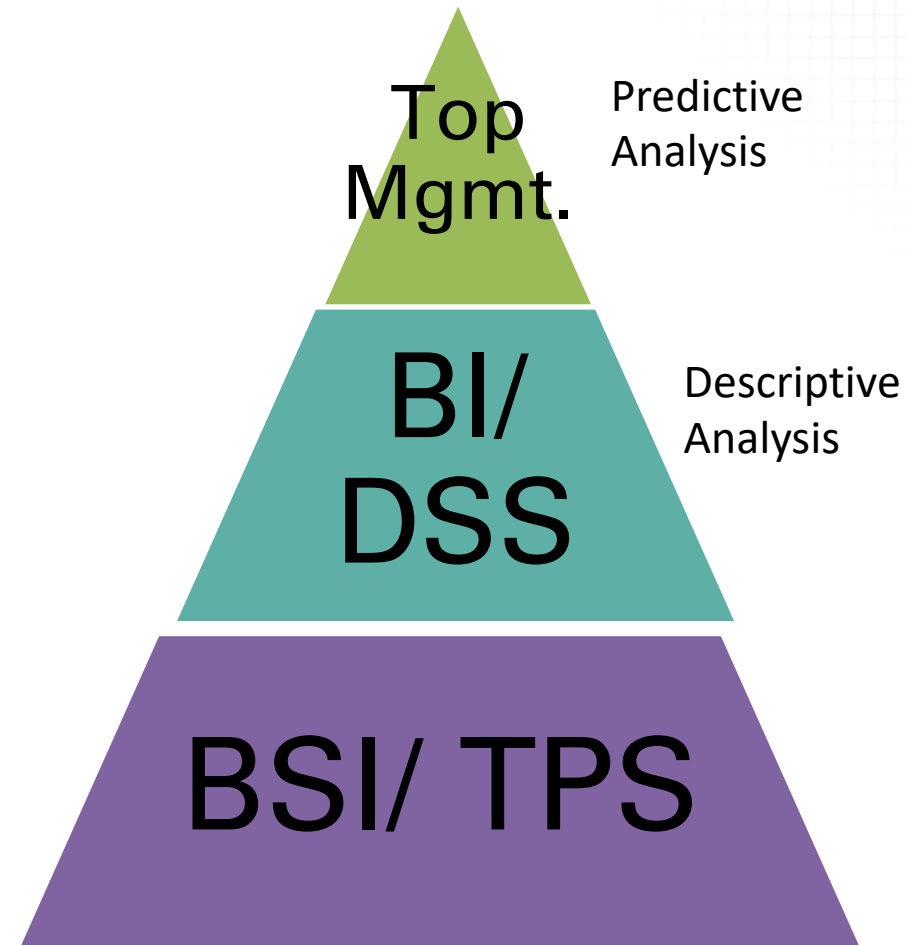
# Business Case

- Problem
  - Inconsistent, exposed data
  - Files or records removed without proper logging authorization and logging
  - Lost data through new deployments, transitions or migrations
  - Change in ownership
- Risk
  - Unusable data
  - Corrupted datasets
  - Damaged files
  - Lose customer trust on data

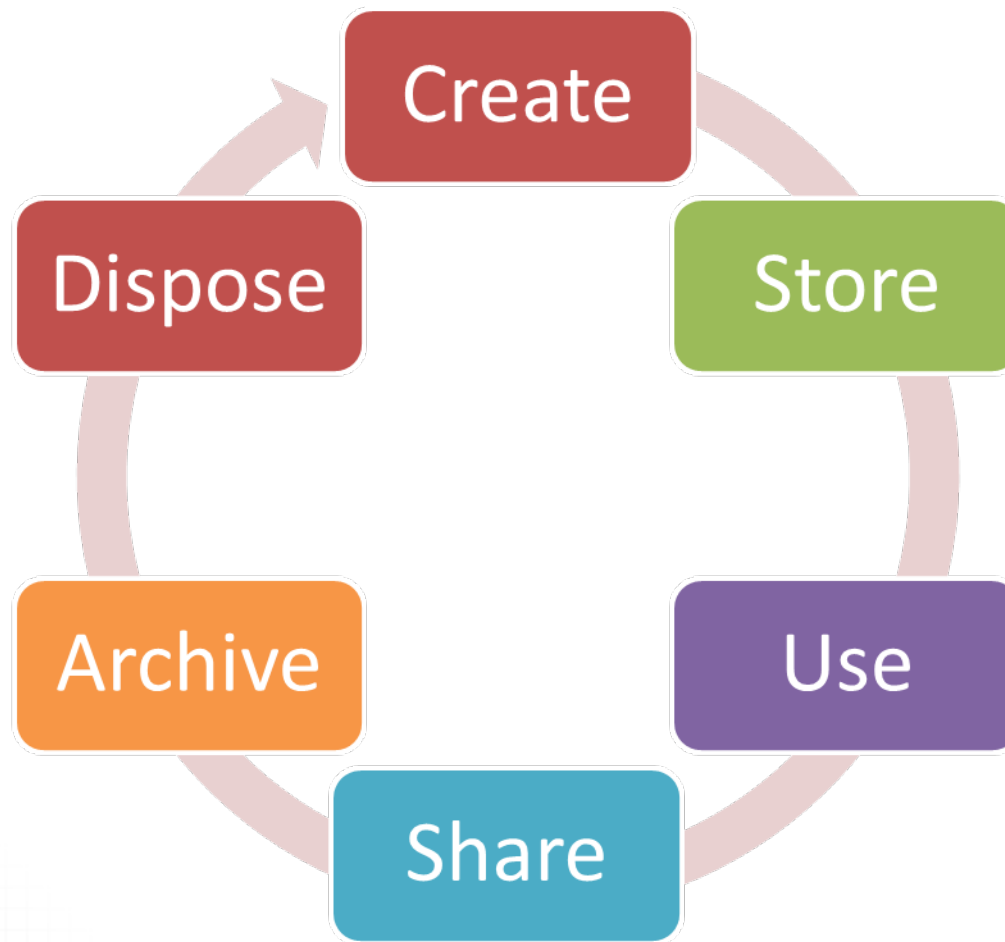


# Supporting Security and Decision Making

- Business value is dependent on quality of data, visibility and accessibility
- Information must be correct, complete, detailed, and integrated to effectively support operations and decision making.



# Data/Information Lifecycle



# Information Lifecycle Management

Create – Data Capture, Data Entry, Input Controls – Due Care



Store – Databases, Files – Management Policies, Data Retention



Use – Processed through processes and workflow



Share – Used and transformed across applications and organizational units



Archive – Dispose –Catalog and send data to secondary storage to free up storage space. Automated or non-automated removal of expired data

## What is Data Governance?

- Set of management activities that organizes, implement and manage a set of structured policies, standards and procedures for effective use of the organization's structured and unstructured data.
  - A sound data governance program includes a governance council, similar to the security steering committee
- Entails policies, standards, procedures, tools and techniques, quality measures, roles, responsibilities and accountability, plus oversee data management operations across the information lifecycle to ensure lean data integrity, availability and confidentiality (CIA) triad.

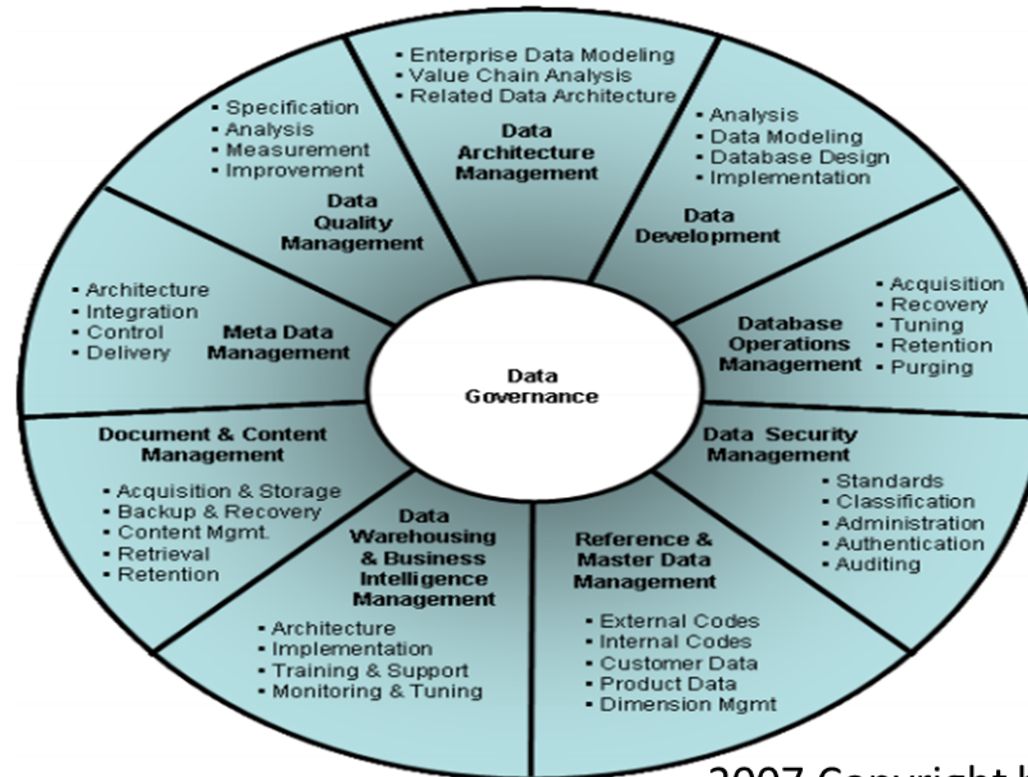
# Data, Information, Knowledge Management

- Knowledge Management is defined by DAMA Intl. (2009) as the “Discipline that fosters organizational learning and the management of intellectual capital as an enterprise resource”.
  - Organizations that recognize the value of data can take concrete, proactive steps to increase the quality of data as a valued asset of an enterprise.





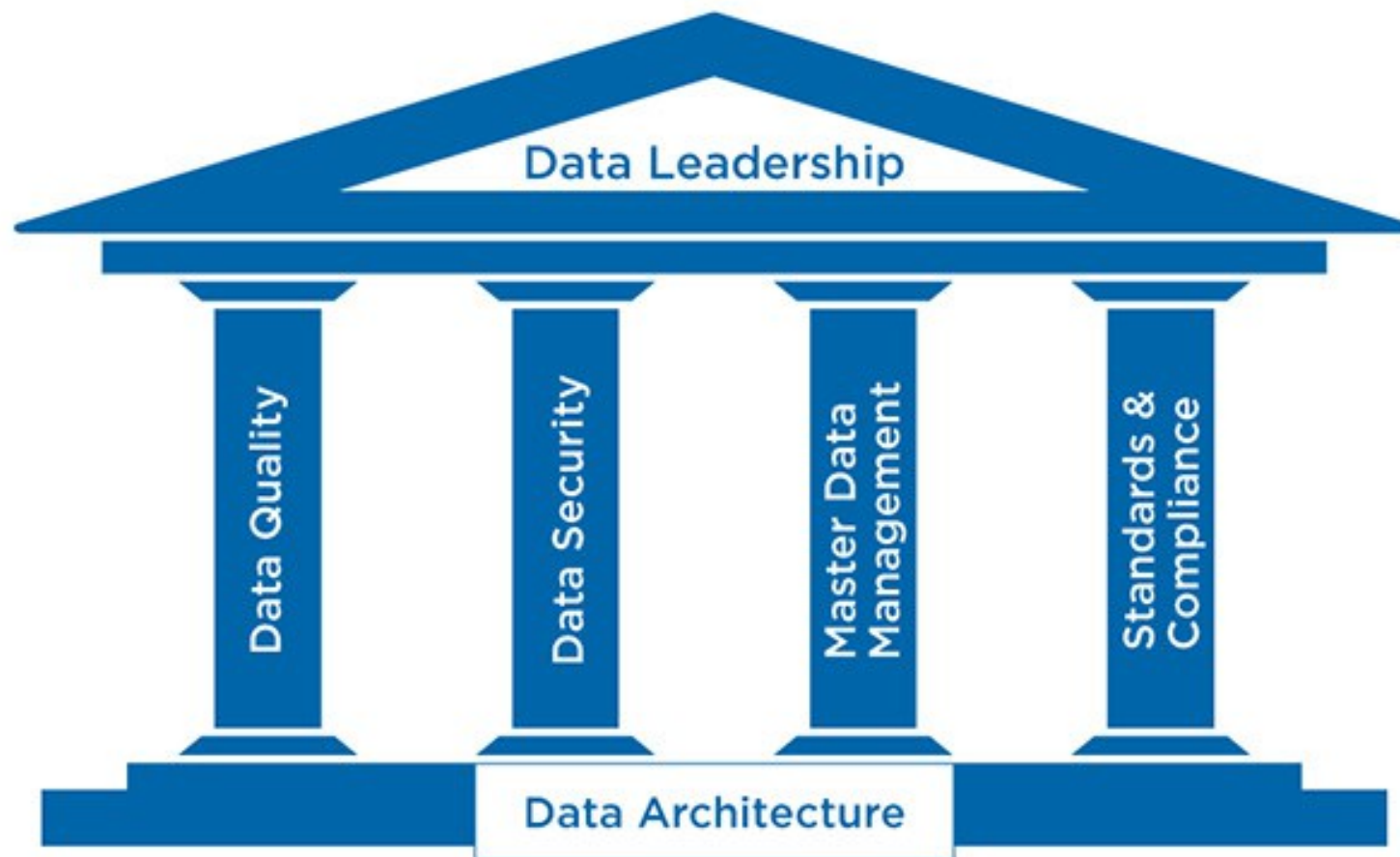
# The DAMA<sup>®</sup> Body of Knowledge<sup>®</sup>



2007 Copyright by DAMA International

Data Management International Organization ([www.dama.org](http://www.dama.org))

# Pillars of Data Governance



## The DAMA<sup>®</sup> Body of Knowledge<sup>®</sup>

- **Data Governance** – Planning, supervision, and control over data management and use.
- **Data Architecture Management** – Defining the blueprint for managing data assets.
- **Data Development-** Analysis, design, implementation, testing, deployment, maintenance.
- **Data Operations Management-** Providing support from data acquisition to disposition.
  - Protection – encryption
  - Media destruction- degaussing, low level formatting, drilling, shredding

## The DAMA<sup>®</sup> Body of Knowledge<sup>®</sup>

- ***Data Security*** – Ensuring privacy, confidentiality, access levels. Involves systems, appliances & processes to protect data and database from intentional or unintentional damage.
- ***Data Quality Management*** – Defining, monitoring, and improving quality, correctness and integrity of data.
- ***Master Data Management*** – Definitions and safeguards of critical data repositories.
- ***Meta Data Management***- Provides information on the identification, data attributes, business and technical definitions of data repositories and data sets.

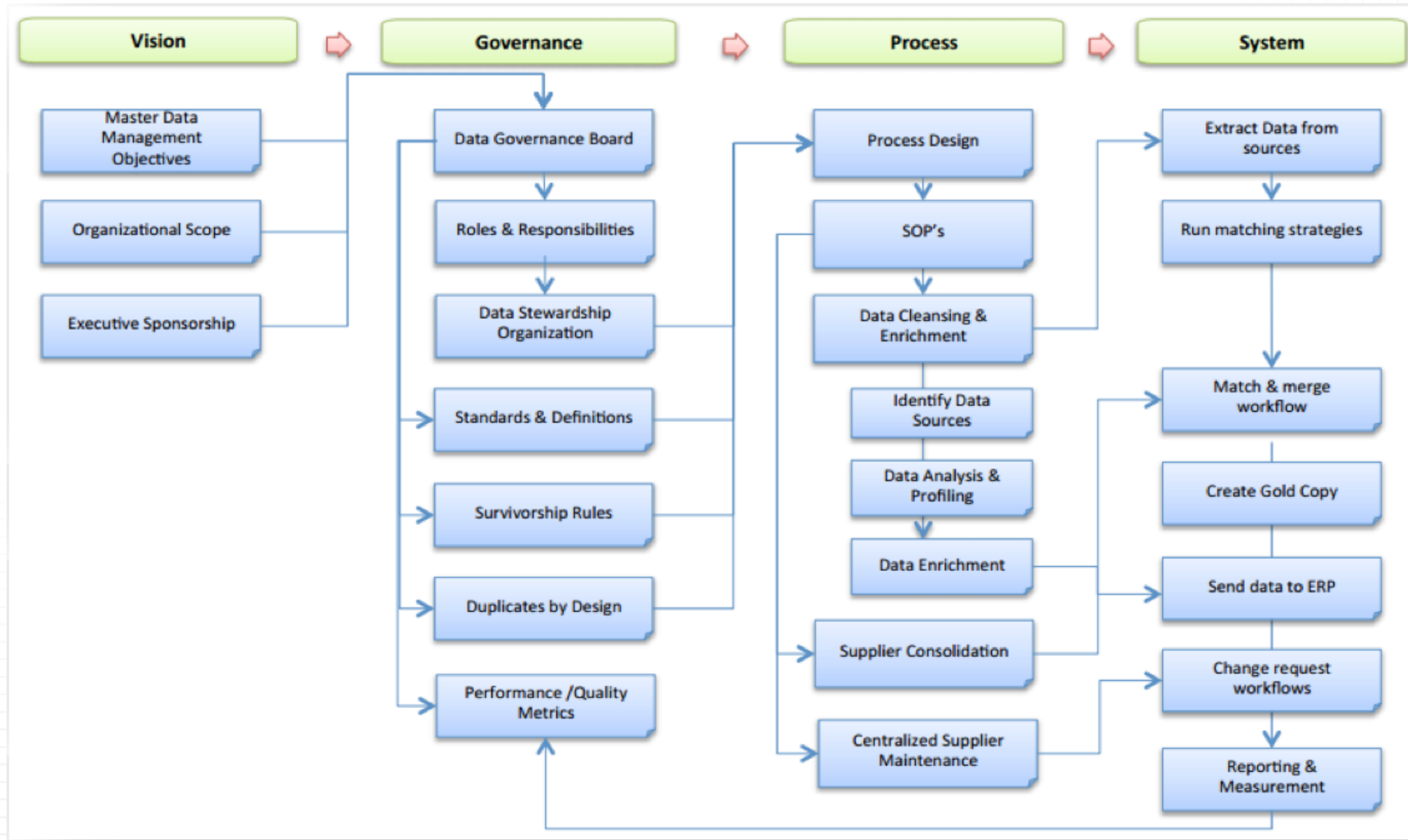


# Master Data Management

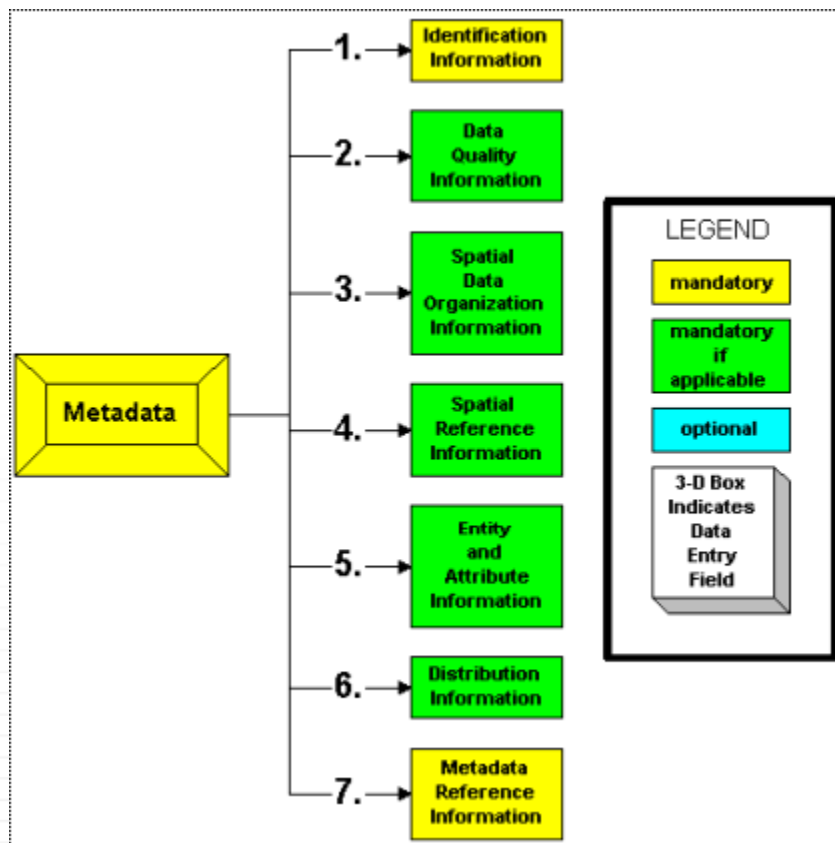


- Master Data Management (MDM) is a comprehensive strategy to determine and build a single, accurate and authoritative source of truth of a company's information assets and deliver this on demand as a service.

# Pillars of Master Data Management



# Metadata Management



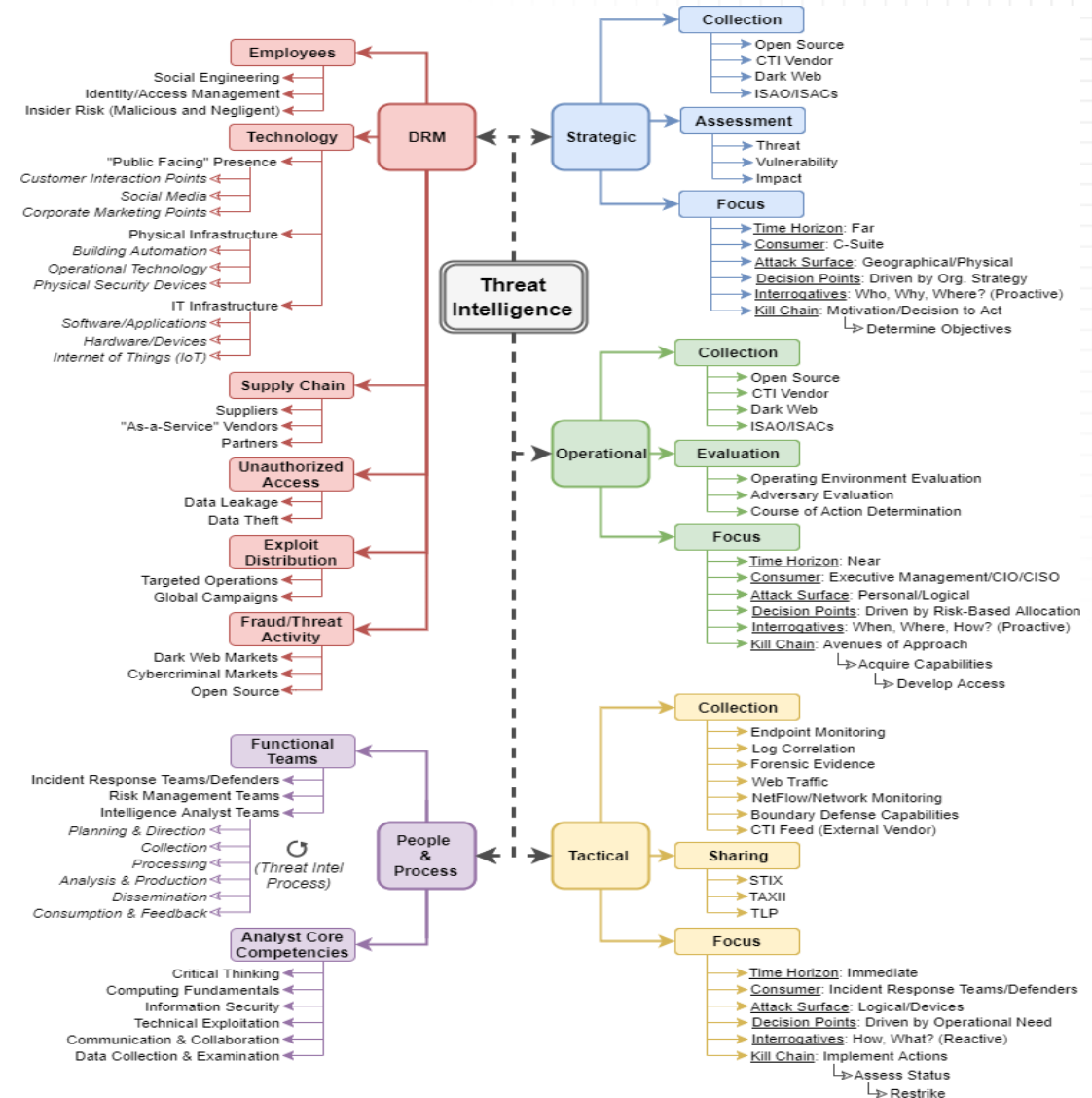
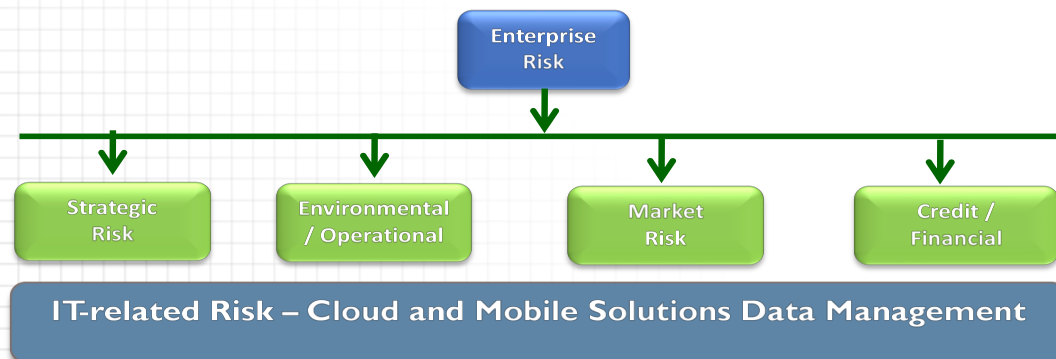
- Metadata management is the oversight of data associated with data assets to ensure that information can be integrated, accessed, shared, linked, analyzed and maintained to best effect across an organization.

# Data Taxonomy

**Taxonomy:** the relationship and importance of risk analysis for information security is stated and justified.

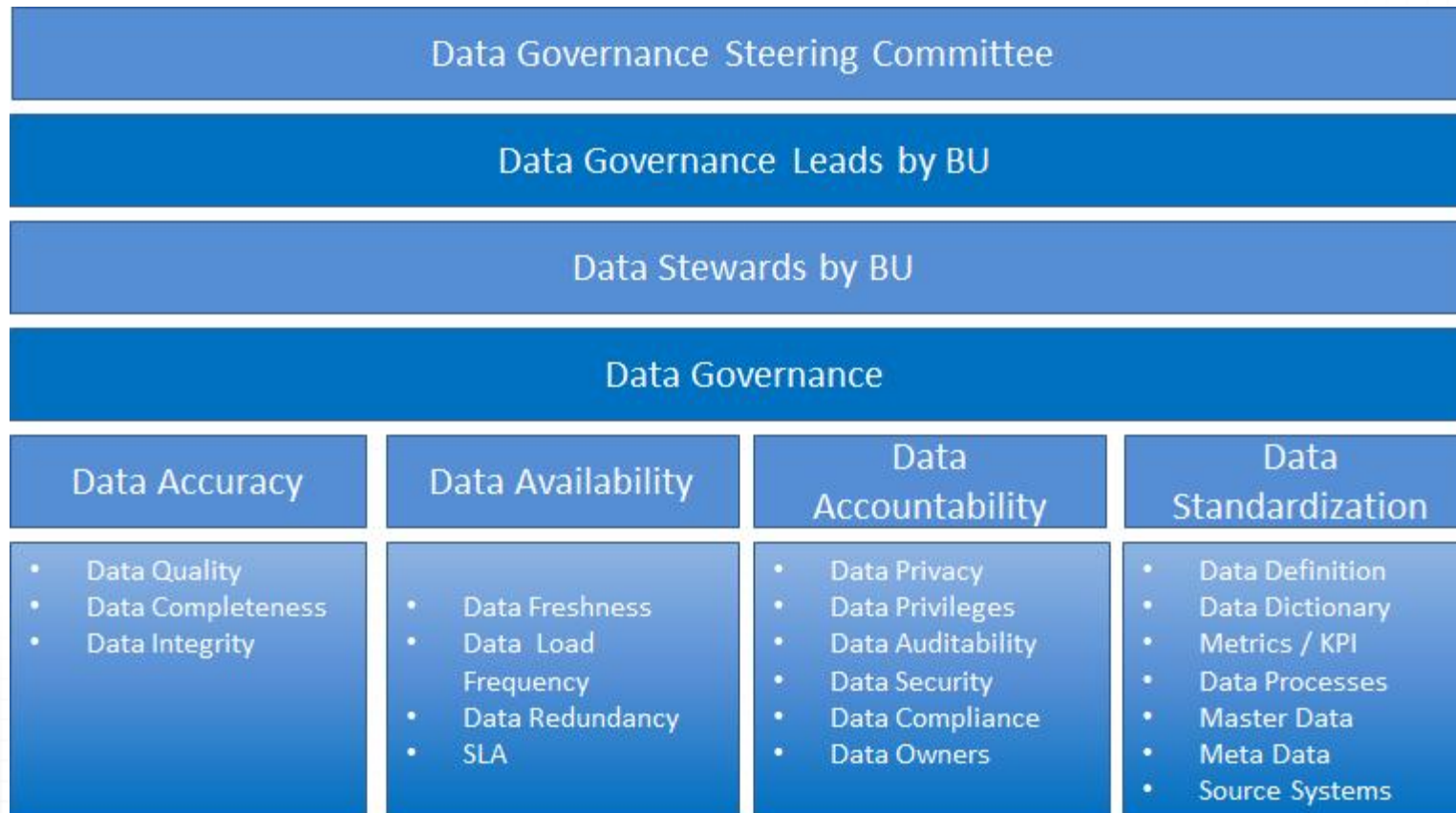
The taxonomy of operational cyber-risks looks to identify, organize, and classify all resources and sources of data and processes that the cyber-risk function must assess and manage—from people issues; hardware-software failures; gaps, faults, or holes in organizational processes; inventory of processes, databases, and applications running as part of the operation or service chain; and external events or activities.

In **data management**, taxonomy represents all the data elements or attributes, and their characteristics and relations within a domain (Knight, 2017).





# Data Governance Council – Security Steering Committee



Source: <https://www.securestate.com/services/steering-committee-meetings>

# Data Governance Council – Security Steering Committee

## STEERING COMMITTEE MEETINGS



Source:

<https://www.securestate.com/services/steering-committee-meetings>

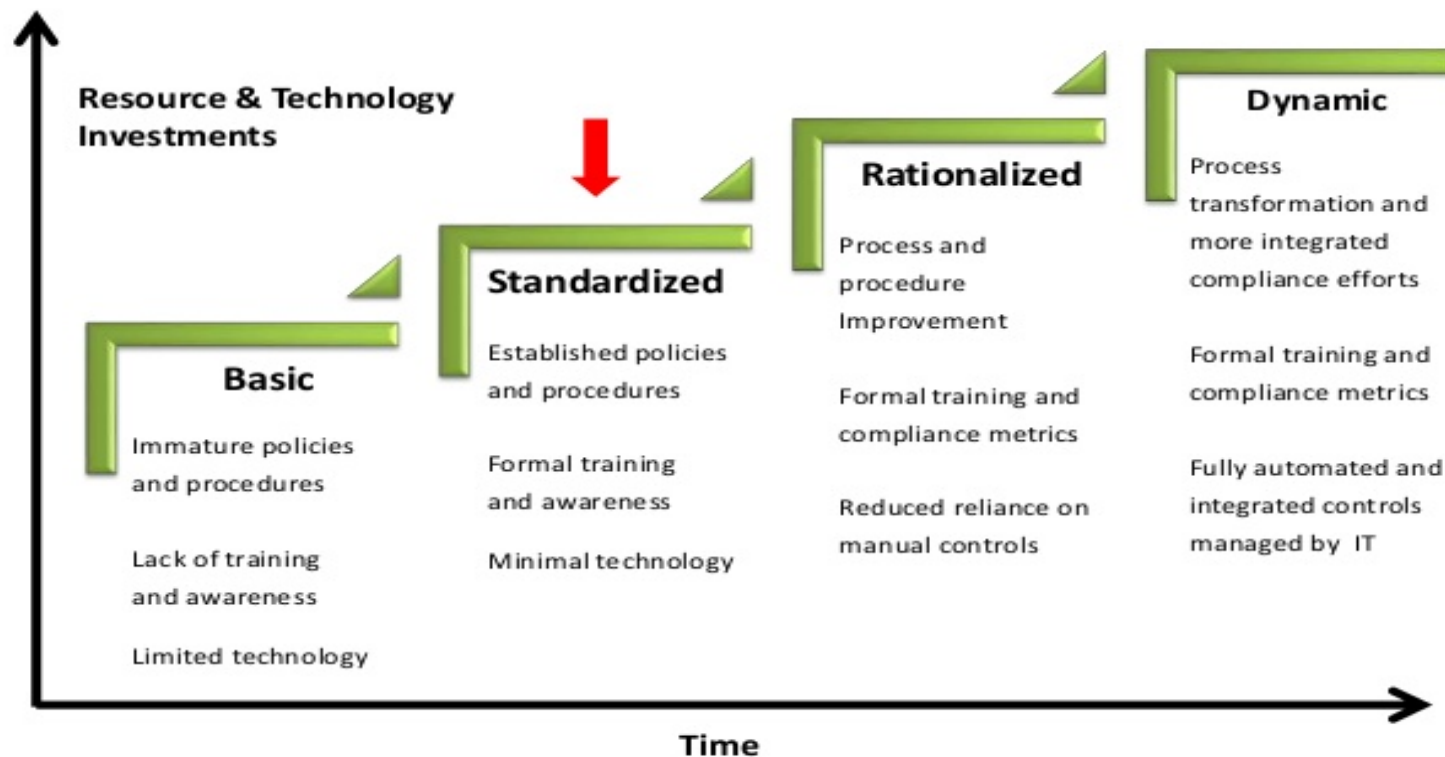


## Steps for Starting a Data Governance Program

- Get top management support
- Set clear goals, purpose and objectives
- Include only domains or governance levels needed based on company size and number of transactions
- Top down/ bottom up approach = design data governance SOP's form top level, but implement them at operational level
- Request Feedback
- Communication

# Maturity Model

## Data Governance Maturity Model







- Anderson, T. (2010). Retrieved from <http://www.guardian.co.uk/cloud-computing/what-is-cloud-computing>
- Berson, A. & Dubov, L. (2011). *Master Data Management and Data Governance*. 2<sup>nd</sup>. Ed. New York. McGraw Hill
- Bitnami, Cloud.com, Zenoss (2010). *2011 Cloud Computing Outlook Survey Results*. Retrieved from <http://cloud.com/cloud-computing-outlook>.
- Clarke, R. (2010). *User Requirements for Cloud Computing Architecture*. Proceedings from the 10<sup>th</sup>. IEEE/ACM International Conference on Cluster, Cloud and Grid Computing.
- Cisco. (2009). *Cloud Computing in Higher Education: A Guide to Evaluation and Adoption*. Retrieved from [www.cisco.com/go/edudatacenter](http://www.cisco.com/go/edudatacenter).
- Enriquez, J. (2011). *The Glory of Big Data*. Popular Science. Vol.279. No.5. pp.31-33
- Hatig, K. (2009). *What is Cloud Computing?*. Retrieved from <http://cloudcomputing.sys-con.com/node/579826>
- ISC2. (2015). *Official Guide to the CISSP CBK*. 4<sup>th</sup>. Ed. New York. CRC Press
- Knight, M. (2017). What is taxonomy? Dataversity. <http://www.dataversity.net/what-is-taxonomy/>
- Mosley M., Bracket, M. Earley, S. Henderson, D. (2009). *DAMA Guide to the Data Management Body of Knowledge. (DMBOK)*. 1<sup>st</sup>. Ed. New Jersey. Technics Publications.
- NIST (2011). *NIST Definition of Cloud Computing. Special Publication 800-145*
- Rouse, M. (2007). *Data Governance*. Retrieved from: <http://searchdatamanagementtarget.com>



SANDRA FONSECA  
PROFESSOR, PROGRAM CHAIR INFORMATION  
TECHNOLOGY  
SCHOOL OF TECHNOLOGY  
NORTHCENTRAL UNIVERSITY  
SFONSECALIND@NCU.EDU





PROJECT MANAGEMENT  
CENTER FOR EXCELLENCE

A.J. CLARK SCHOOL OF ENGINEERING  
Civil & Environmental Engineering Department

UMD Project Management Symposium  
May 7-8, 2020

Slide 40





# Template Basics

- Install template fonts (Univers)
- Use template layouts
- When pasting content, use destination formatting / text only



# Dos and Don'ts

## PRESENTATION BEST PRACTICES



# Do...

- Practice!
- Stay < 18-20 min to allow for Q&As
- Stick to template styles
- Cite all sources



# Don't use

- Inconsistent styles
- Cramped or low resolution content
- Poor English
- Misinformation



# Questions?

Contact Kathleen Frankle

[kfrankle@umd.edu](mailto:kfrankle@umd.edu)

301-405-8271