North Highland
Data and Analytics

Data Governance Considerations for Big Data Analytics

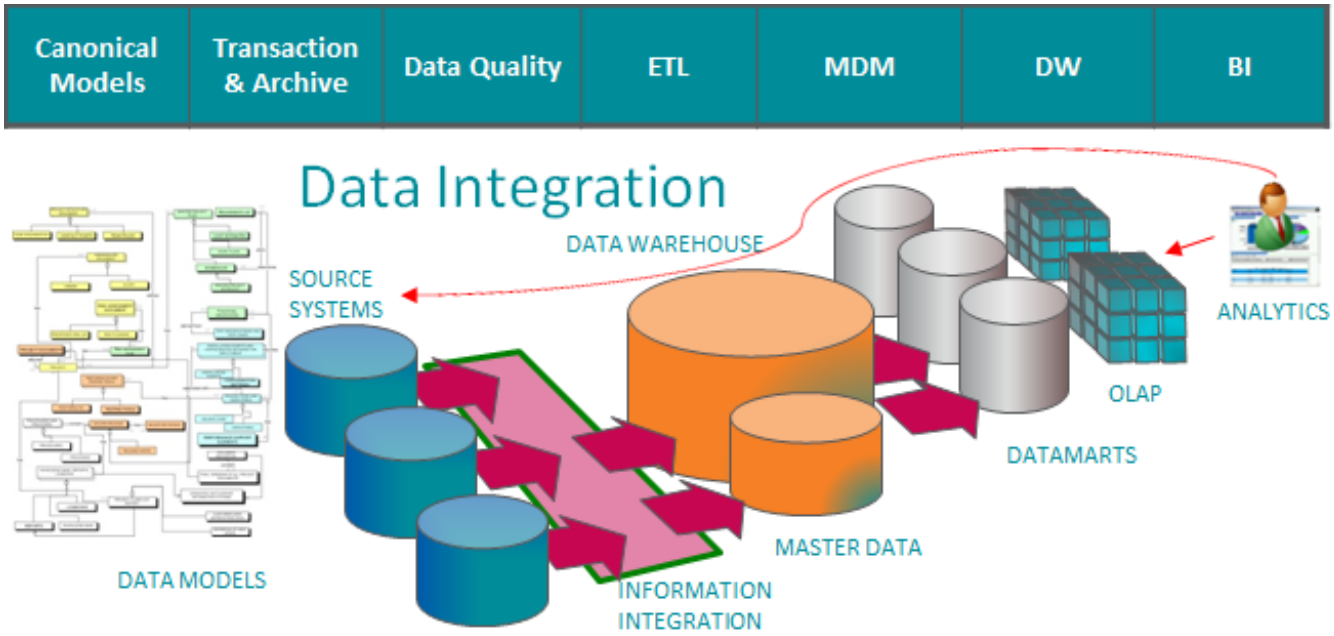
northhighland.
WORLDWIDE CONSULTING

Data and Analytics

Agenda

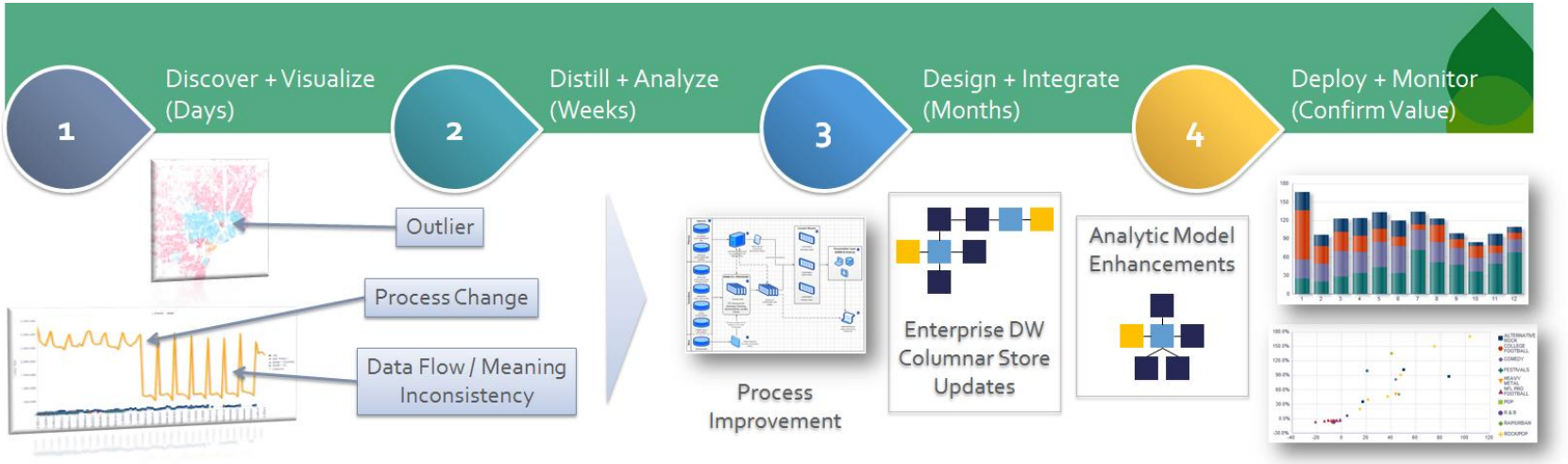
- Traditional BI/Analytics vs. Big Data Analytics
- Types of Data Requiring Governance
- Key Considerations
- Information Management Framework
- Organizational Framework
- Conclusion – Wrap Up!
- Questions

Traditional BI/Analytics



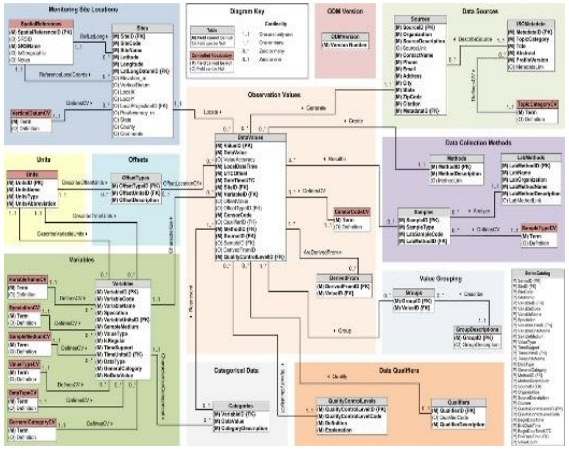
- Canonical models provide fixed reference points.
- Transactional events generate archive-worthy data and business and IT processes assure data quality.
- Integration approaches such as Extract Transform Load (ETL), Enterprise Information Integration (EII), and/or Enterprise Application Integration (EAI) are used to unify data.
- Master Data Management helps accurate matching and consistent aggregation.
- Data warehouses (real or “virtual”) act as ultimate repository for supporting BI.

Big Data Analytics



- Big Data technology allows us to capture and process massive amounts of data at the speed of business
- New ways to manage unstructured and semi-structured information previously unexplored
- The highest value comes from making connections across data sources, types, functions, etc. versus standalone environments / applications
- Traditional BI/Analytics informs of past activities and performance
- Advanced Analytics is typically future-oriented and leverages some type of statistics and algorithms to present a Simulation, Descriptive Analytics or Predictive Analysis

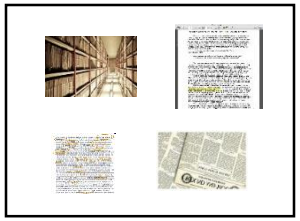
Types of Data Requiring Governance



Structured Data

"Web & Social Media"

Highway Traffic Flow



Legacy Documents

Images & Video

Traditional BI/Analytics

IT typically provides information via extracting data from a data warehouse and presents to the business in pre-defined reports.

Big Data Analytics

Analytical platform is provided that lends itself to be more variable/flexible to analyze data from multiple and varying types of sources including Traditional.

Key Considerations

- Analytics and Big Data Challenges
  - Its vast, variable and high velocity. Most governance programs and models depend on structured data delivered in predictable volumes.
  - Its not just numbers and names, its video, photos, text, tweets, posts, machine data, etc.
  - Exploratory analytics (research) depends on new or blended data sources – governance rules get in the way.
- Where you should focus attention
  - What are the analyst segments (audiences) you need to serve. Research analysts need raw data, financial analysts probably need more governed data.
  - Define the level of governance needed for each audience and data source (Ex: Twitter: meta-data around hashtags so you can mine). Publish a 3-5 step standard based on Raw to Gold (gold being the top standard everyone can trust).
  - Some data sources shouldn't be managed (free, open) BUT you shouldn't certify reports used with those data sets.
  - PII tokenization schemes are important to allow exposure of data to analysts while still securing data.
  - Look out for new trends in automatic data discovery and tagging to allow automated governance.

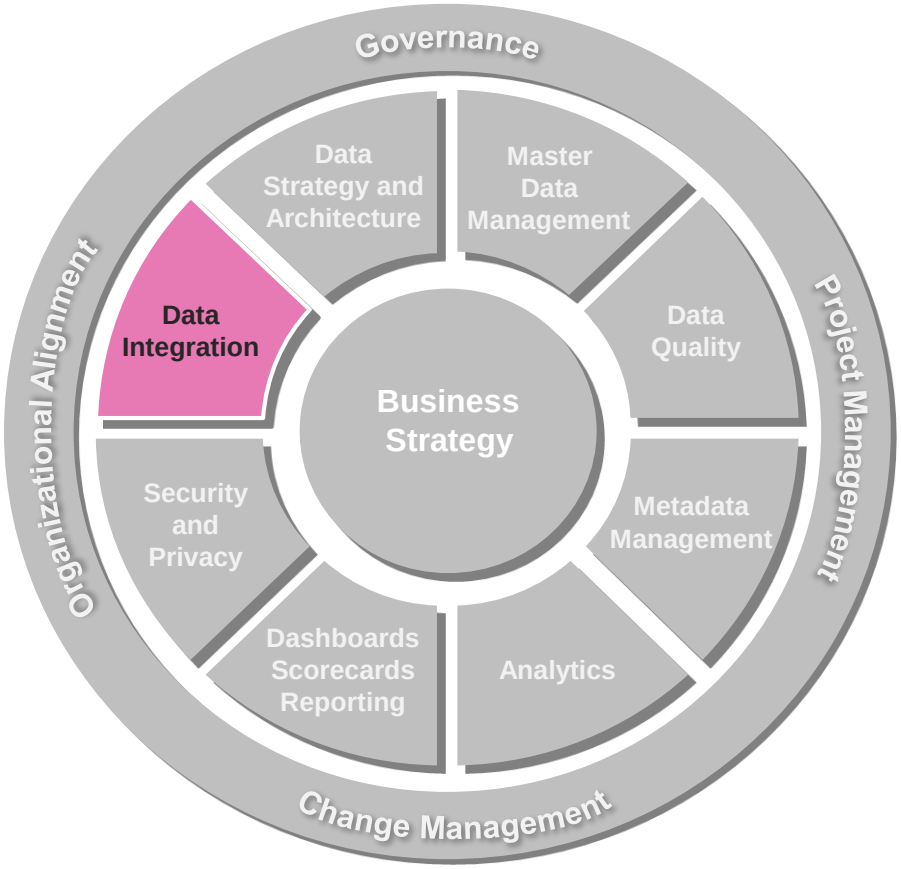
Information Management Framework

Data Governance is critical for Advanced Analytics model implementation but not necessarily exploratory analytics.

A solid Information Management Framework is critical to help establish confidence in the data that will be used in Advanced Analytics models that are put into production.

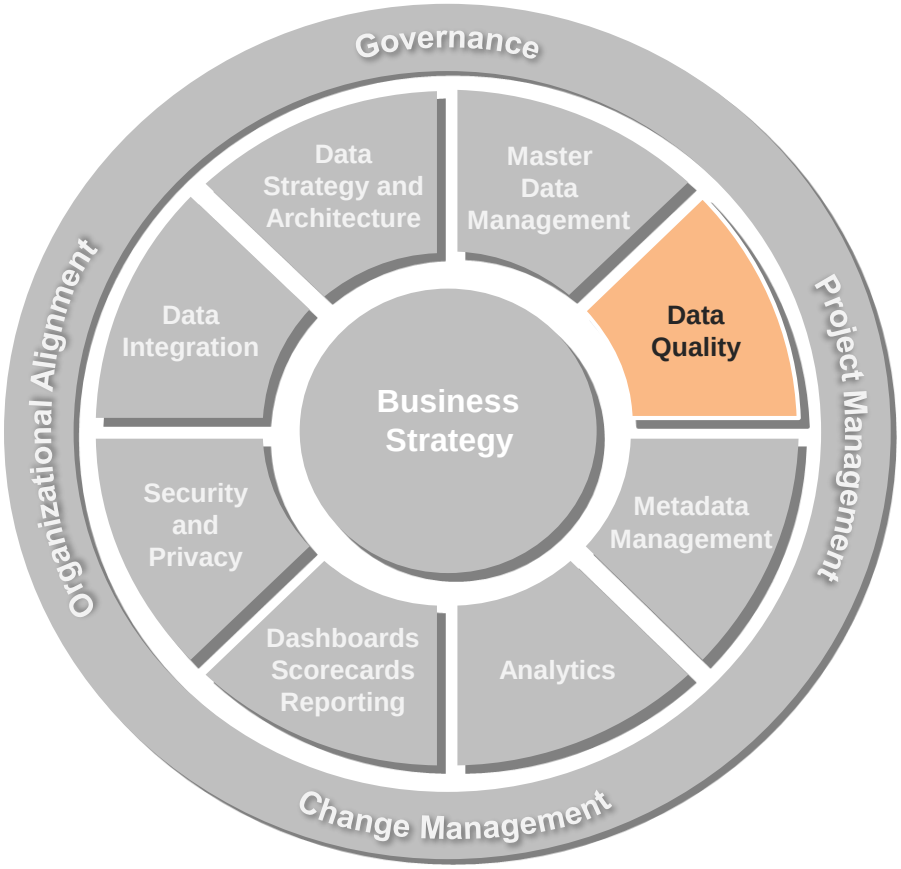
Data Integration

- To get a full 360-degree view of a customer or product, careful attention must be given to how external data is integrated into the data warehouse.



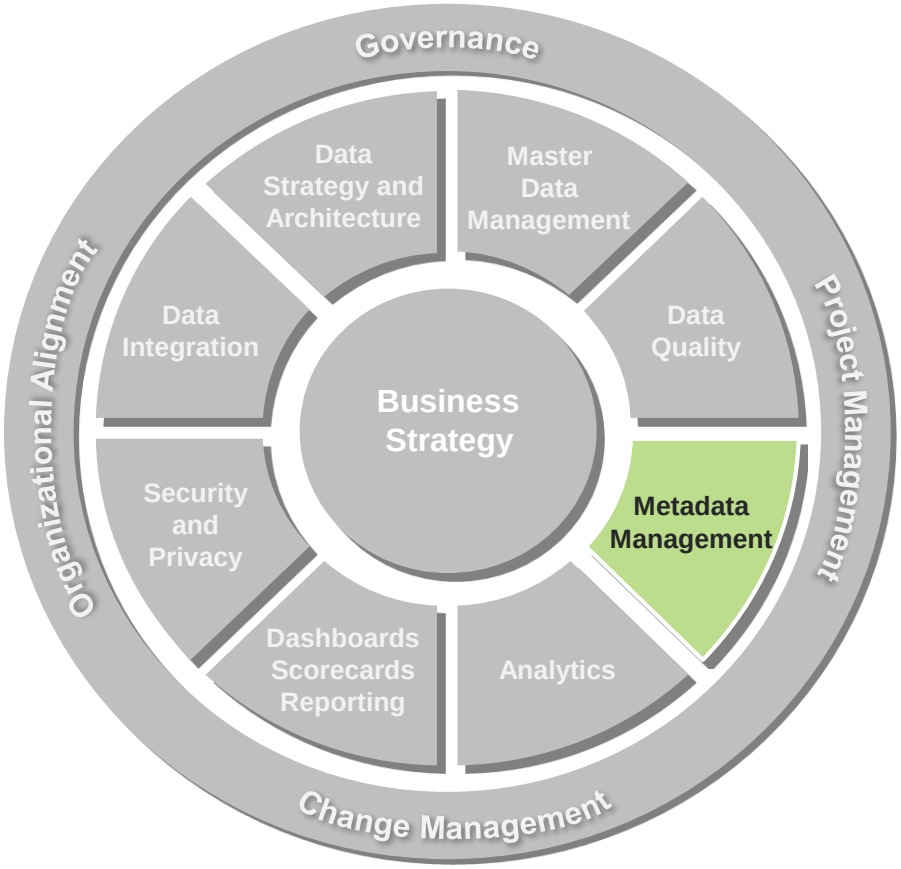
Data Quality – Data Cleansing

- Too much cleansing of data may be harmful to Advanced Analytics.
- Be careful not to overdo it and remove some of the analytic value that would otherwise be found in the data.
- Consider staging data in 3 categories: Raw, Basic and Gold Standard.



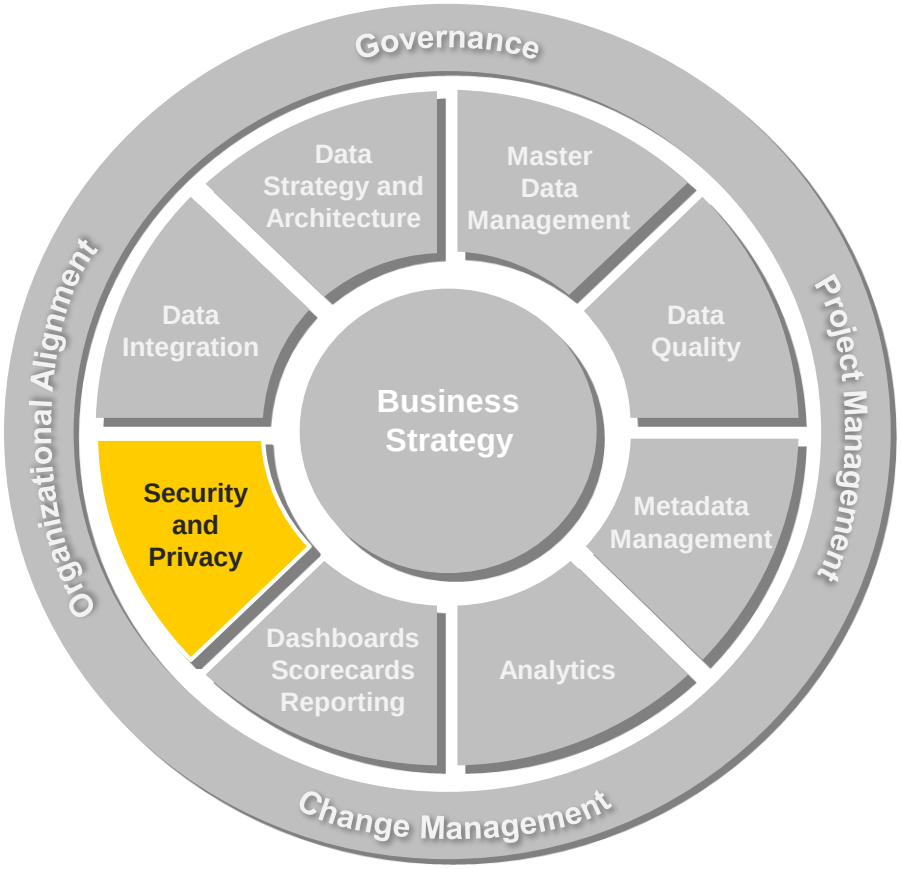
Metadata Management

- It is critical to know the meaning of the data regardless of the source
- Make it accessible to both business and IT
- Track lineage from the source to analytics.
- Use common definitions in order to help establish consistent conclusions.



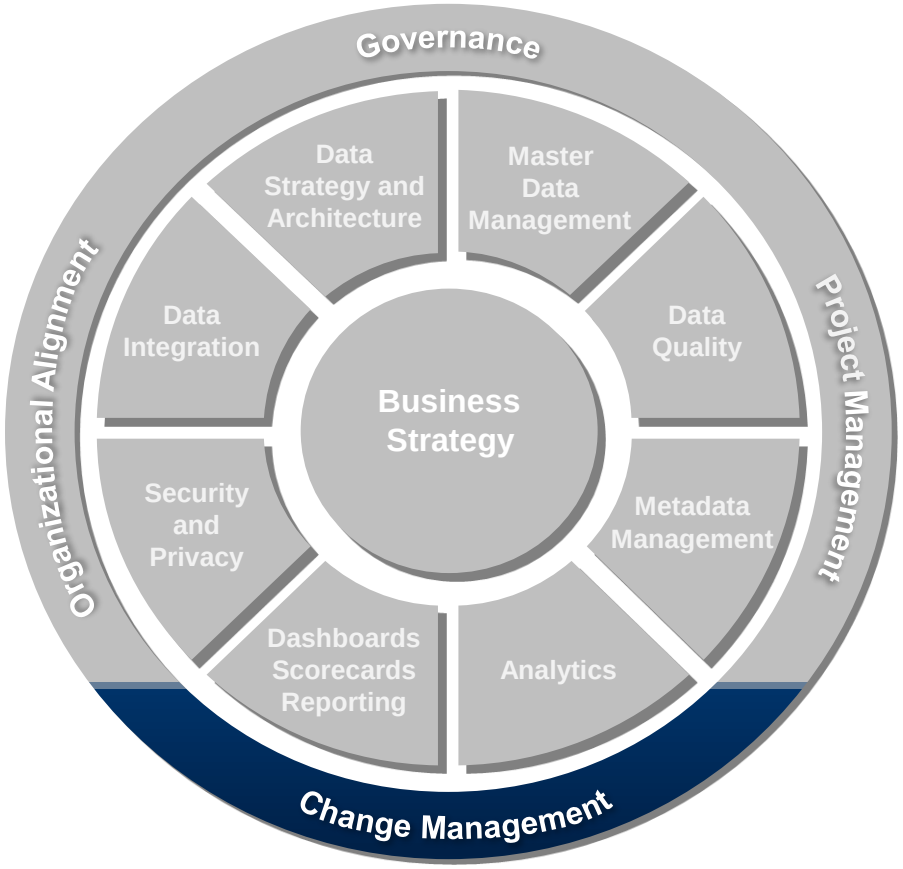
Security & Privacy

Understand what portions of the Big Data being collected should be maintained as private or sensitive to meet company compliance.



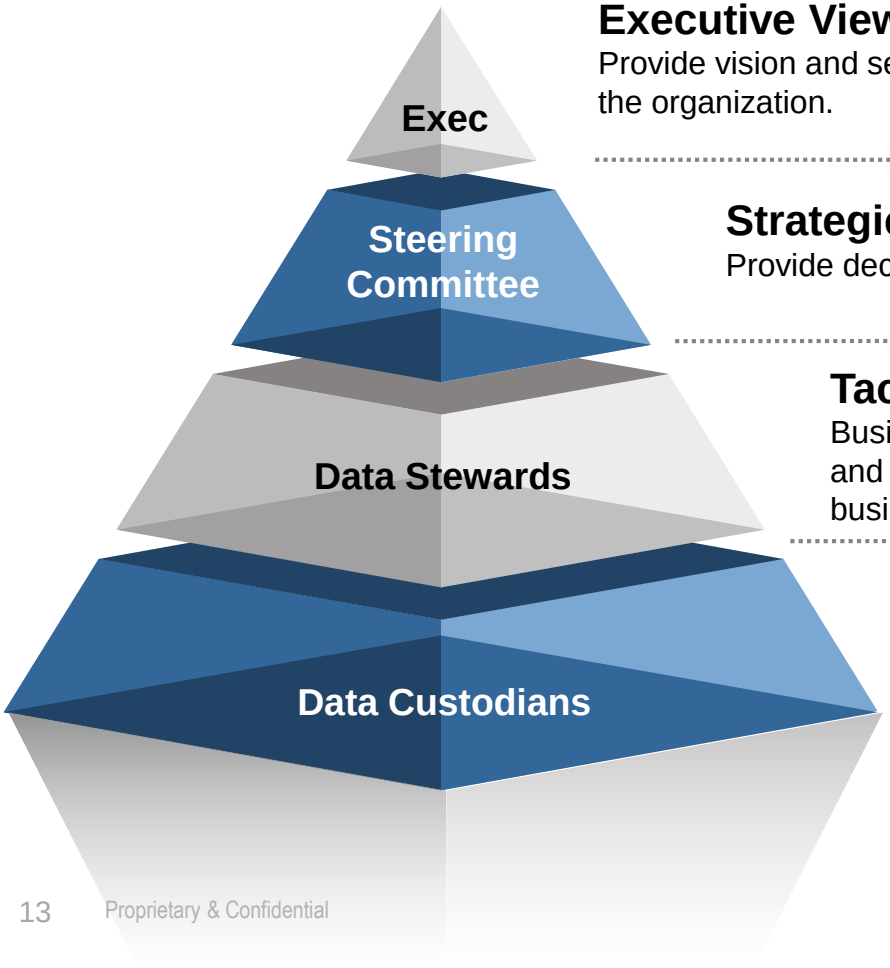
Change Management (Training)

Conduct sufficient training throughout the organization to help executives, managers, etc. grasp the fundamentals of Advanced Analytics so they can ask the right questions of their data analysis teams.



Organizational Framework

Each level of data governance has a specific role and view of the overall Data Governance program. It is important to understand how impactful and critical each level of involvement is to attaining the overall objectives.



Executive View (Sponsorship)
Provide vision and set strategic priorities. Provide executive voice into the organization.

Strategic View
Provide decision making and oversight. Set priorities and goals.

Tactical View
Business users with expert knowledge of business processes and how data is used. Typically the “go to” person within their business group for all data-related questions.

Operational View
Focus on the underlying infrastructure (technical environment, databases, etc.) and activities required to keep the data intact and available to users.

Conclusion – Wrap Up!

Today, analytics systems are still somewhat brittle and complex.

We will continue to see a trend for making them easier through self discovery, packaged algorithms, packaged purpose built predictive analytics, etc.

We also see a trend of seeing more analytical tools embedded inside traditional visualization tools, accessing larger volumes of structured and unstructured data.

Therefore, standard approaches to data governance must be tailored for Advanced Analytics of the future.

Getting these new approaches in place is critical to build confidence in the analytics being performed on ever increasing volumes of information.

Questions?

northhighland.
WORLDWIDE CONSULTING

Data and
Analytics

Thank You

Dan Montgomery

Director, Data and Analytics

Dan.Montgomery@northhighland.com

https://www.linkedin.com/pub/dan-montgomery/0/149/4b2