## Session 1: Data Governance Overview

* + The difference between Data Management & Data Governance
    - Hint: Data Management deals with the structure and performance of the data. Data Governance deals with accessibility and quality of the content of the data
  + What is data, and why should it be governed
    - An asset that is non-consumable, and retains value
  + A PROGRAM, not a PROJECT
    - You can stop it when you stop your HR program or Supply Chain program
    - Justify it with adverse events
  + If data is an asset, then the asset needs to be managed and maintained
    - Use the “Truck Fleet Analogy”
  + Who are the actors?
    - Data Owners, Data Stewards, Application Stewards, Analytic Stewards
  + What are the models?
    - Depends on the goal.
      * If you’re trying to keep your warehouse clean, consider a central model
      * If you want better data throughout the org, consider a distributed model
  + Introduce the 6 DG activities

# Session 2: Data Domain Governance

* + What is a domain?
    - Many disciplines use the term “Domain”. What does it mean to DG?
  + Why is Data Domain the first activity?
  + The Domain selection process
    - Find an easy one
    - Find a long-term one
    - Find others based upon enterprise strategies and important KPI’s
  + Find the Data Owners
    - This might be tricky. A Data Owner needs respect and authority across the organization.
  + Find the Stewards
    - This might be easy. They already exist in your organization, and are recognized in that capacity. Their role is just not formalized
    - We’re asking people who spend X hours per week figuring out what happened to spend <X hours per week making sure nothing happens

# Session 3: Data policy and Strategy

* + What is a Data Strategy?
    - A set of guidelines and patterns for decision making
  + Your Strategy needs to be flexible to changes in business need, and be able to take advantage of new technology
  + Data Strategy must follow Analytics Strategy
  + Access policies must follow
    - Balance the HIPAA “minimum needed” with the organization’s need for insight, and opportunity for monetization

# Session 4: Metadata Management

* + Easy definition: “The stuff you need to know to make the best use of your data”
  + The most esoteric concept to sell to Management
  + Critical to democratizing data
    - It’s the key to taking the “tribal knowledge” in your Analytics areas, and making it available to all
    - Highly participatory
  + You will get resistance from people who believe they are the only ones who know enough to access the data.
  + The three levels of Metadata
    - Dictionary, Catalog, Glossary
  + The magic of Data Lineage
    - I like the term “data pedigree” – do you really want to make decisions from “mongrel data”

# Session 5: Master and Reference Data Management

* + The second most esoteric concept
  + Initially high investment, subsequent high payoff
  + Can greatly shorten time-to-insight
  + What is Master Data?
    - The 5 P’s – Patients (or People), Providers, Places, Payers, Procedures
    - Contrast Master Data in Healthcare versus other industries
  + What is Reference Data?
    - The obvious – ICD, CPT, SNOMED
    - The not-so-obvious – internal codes – encounter types, financial classes,

# Session 6: Data Quality Management

* + Initially low investment; subsequent high payoff
  + Find the adverse events, and springboard off of those
  + Data profiling – where it all starts
    - How many duplicate keys do you have? How many different gender codes do you have?
  + Move to more advanced quality analyses as you know more about the data
  + The types of data quality issues
    - Validity, Completeness, Accuracy