tamr
Leverage **All Data**

My background

Math, Stats, CS / ML

Data Science, Data Engineer

Solution Architect, Presales

Entity Resolution and Data Quality Problem
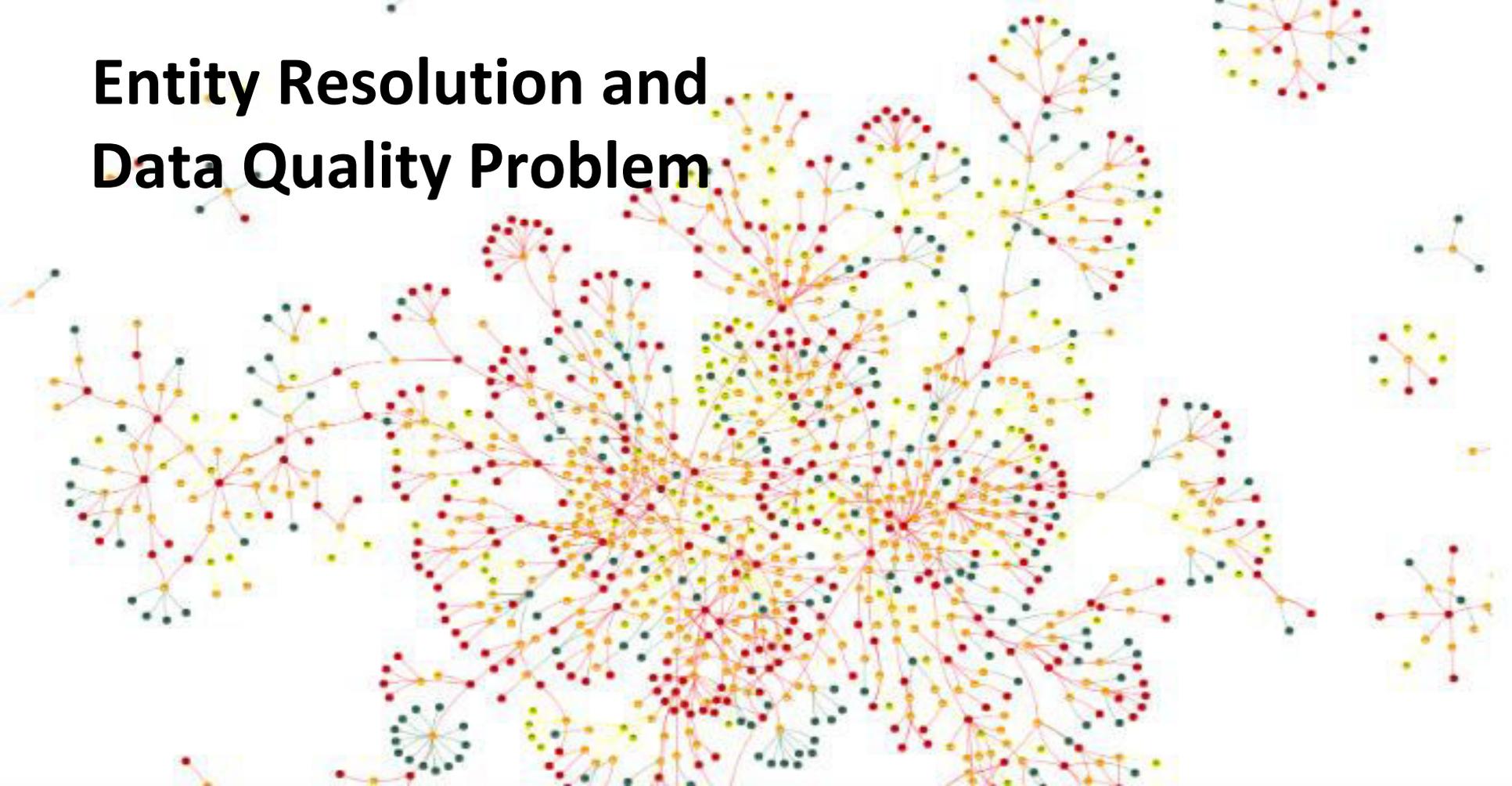
Traditional Methods Don't Scale with Number of Sources

Marginal Cost

- ETL + Data Warehouse
- MDM Tools
- Professional Services

Opportunity For Strategic Data Asset

Number of Siloed Data Sources

tamr
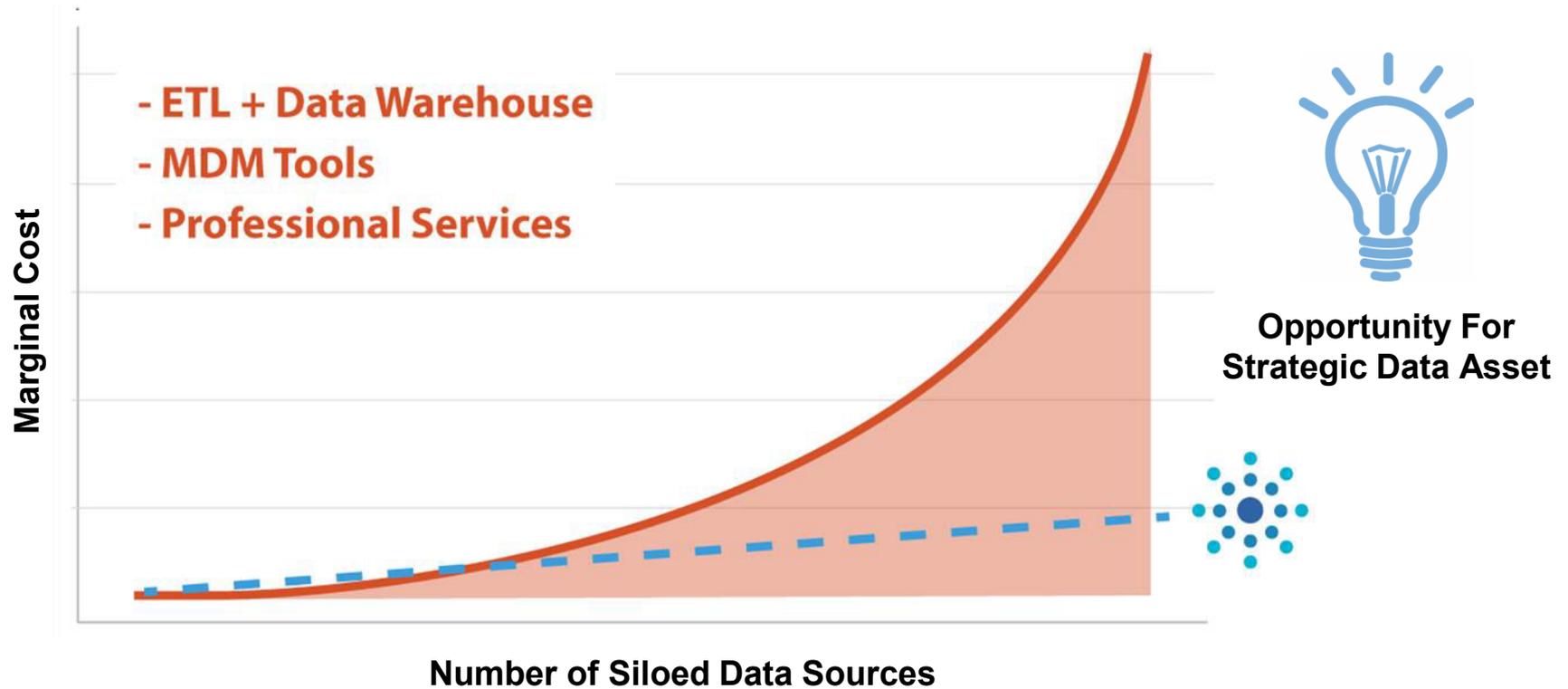
Why I get excited about Enterprise: Scale, Scale, Scale

Problem

8
Divisions

500+
ERP Systems

100K+
Suppliers

10M+
Customers

Solution

Suppliers

Customers People Logistics

Parts

Results

Sales Revenue
Marketing Impact

Total
Landed Cost
$300M+
0.5% of Direct Spend
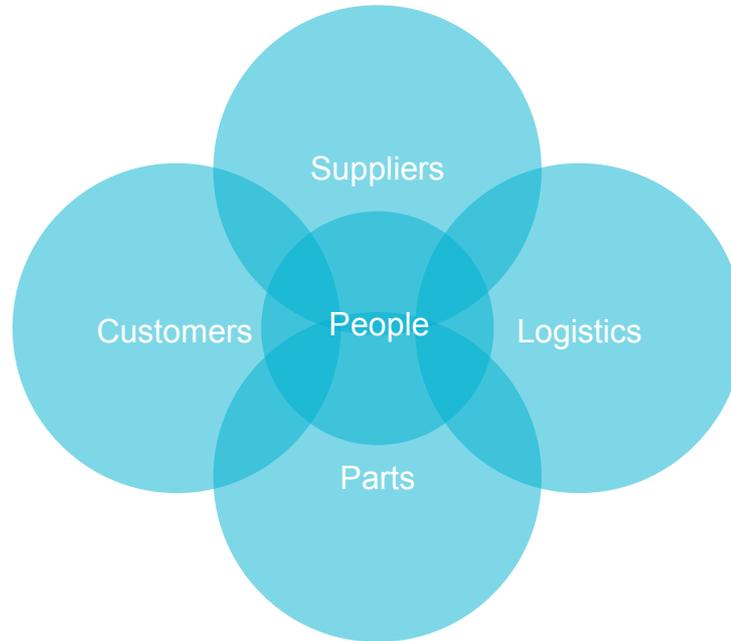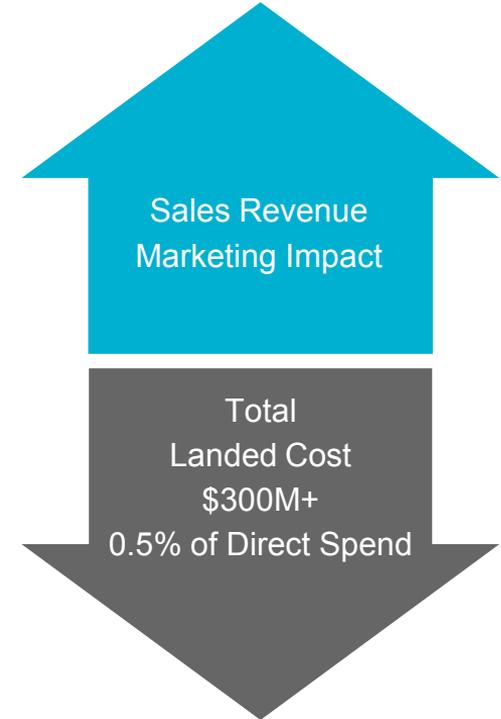
tamr

Dirty Little Secret: Data Variety in Enterprise

What most people think enterprise data looks like
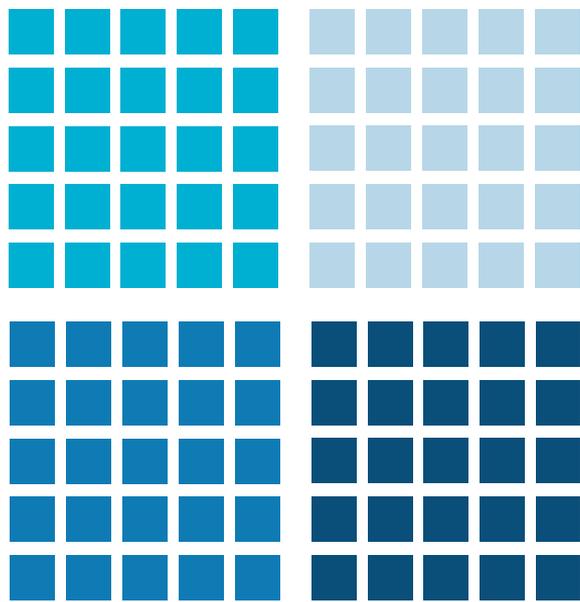
What enterprise data is really like - "random data salad"

Prone to constant change/entropy

M&A

Politics

Dynamic Schema DBs - Mongo et al

Legacy Burden

Restructuring

Leadership Changes

"Data Hoarding"
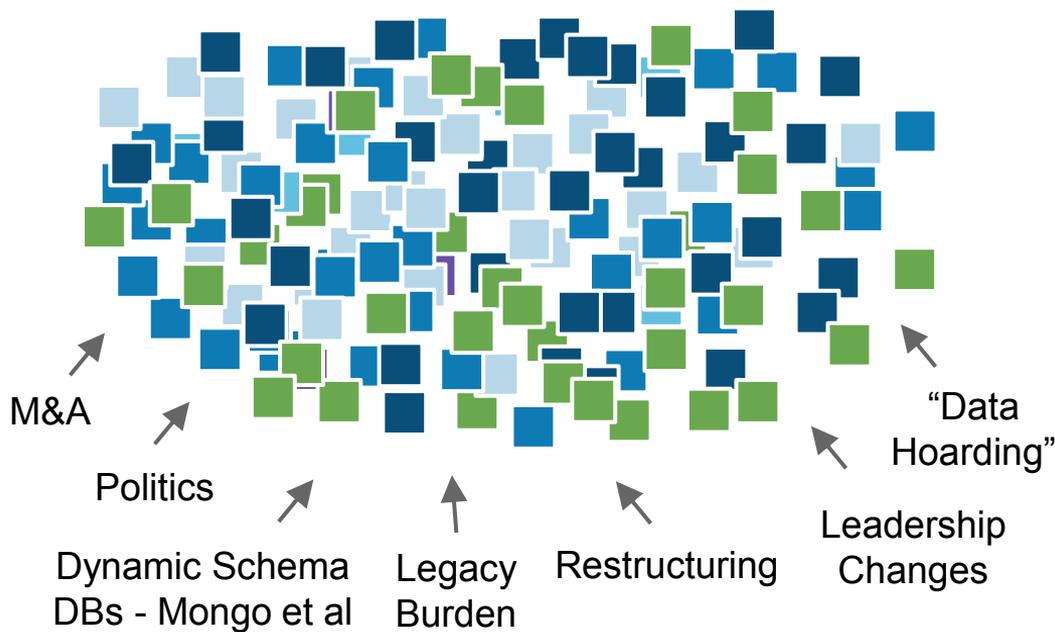
What Tamr Does

Tamr solves the enterprise data variety problem to power transformative analytic and operational outcomes.

10X Reduction
In New Data Set Integration
From 6 Months to 2 Weeks
▷ Video Case Study

$500M+ Savings
From Sourcing Analytics
Across Businesses
▷ Video Case Study

Customer Insights
Unified buyer profiles across
siloed dealer systems in 30+ geos
▷ Video Case Study

5000+ Studies
Unified clinical study data
to empower researchers
▷ Case Study

tamr

Reality for Global Corporate IT as Data Broker

Most data is untreated + unprepared for expensive analytics tools

Finance

HR

Marketing

Sales

Divisions

Manufacturing

Engineering

tamr
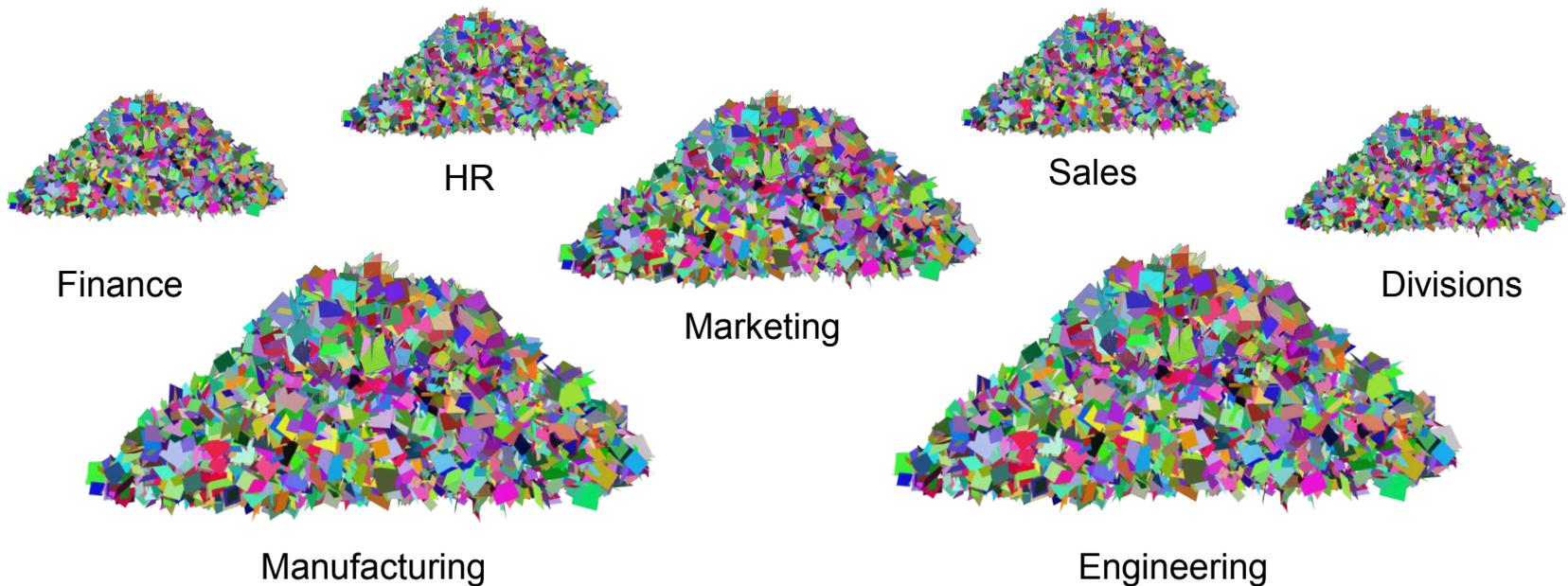
Some Options

Option #1 - <u>Deny Variety</u> - use information that is easiest/closest

Option #2 - <u>Manage Variety incrementally</u> - using traditional approaches:

- Standardization
- Aggregation
- Master Data Management
- Rationalize Systems
- Throw Bodies at it
- Improve Individual Productivity

Option #3 - <u>Embrace Variety</u> using probabilistic/model based approach - Tamr

tamr

Option #1: "Deny" Variety

Use only the information that is closest, most familiar, easiest to obtain
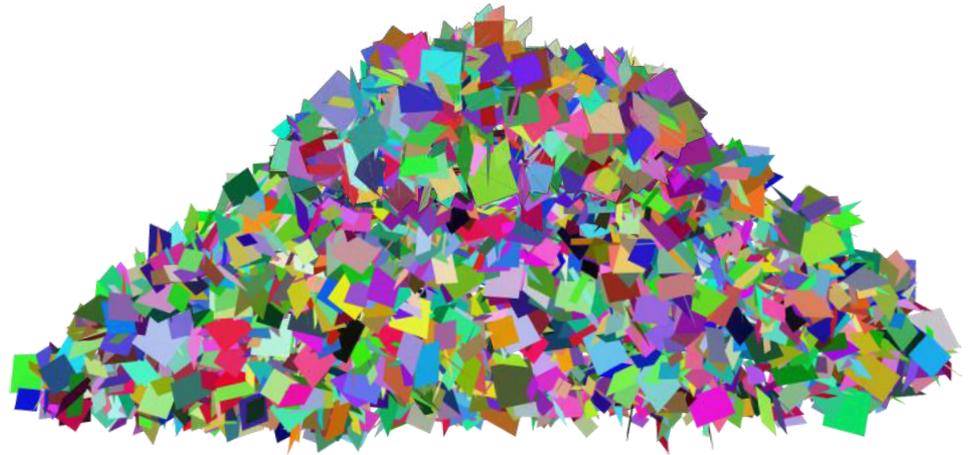
Option #2: "Manage" Variety Using Traditional Approaches

Traditional Data Management Approaches: Necessary but not sufficient

- **Standardization**
- Aggregation
- Master Data Management
- Rationalize Systems
- Throw Bodies at it
- Improve Individual Productivity

One Schema to Rule them All

Option #2: "Manage" Variety Using Traditional Approaches

Traditional Data Management Approaches: Necessary but not sufficient

- Standardization

- **Aggregation**

- Master Data Management

- Rationalize Systems

- Throw Bodies at it

- Improve Individual Productivity

Option #2: "Manage" Variety Using Traditional Approaches

Traditional Data Management Approaches: Necessary but not sufficient

- Standardization
- Aggregation
- **Master Data Management**
- Rationalize Systems
- Throw Bodies at it
- Improve Individual Productivity

Option #2: "Manage" Variety Using Traditional Approaches

Traditional Data Management Approaches: Necessary but not sufficient

- Standardization
- Aggregation
- Master Data Management
- **Rationalize Systems**
- Throw Bodies at it
- Improve Individual Productivity

Option #2: "Manage" Variety Using Traditional Approaches

Traditional Data Management Approaches: Necessary but not sufficient

- Standardization
- Aggregation
- Master Data Management
- Rationalize Systems
- **Throw Bodies at it**
- Improve Individual Productivity

Option #2: "Manage" Variety Using Traditional Approaches

Traditional Data Management Approaches: Necessary but not sufficient

- Standardization

- Aggregation

- Master Data Management

- Rationalize Systems

- Throw Bodies at it

- **Improve Individual Productivity**

Logical Evolution to Probabilistic/Model-Based Approach

Probabilistic (Tamr) *Complements*, NOT Replaces, Deterministic (MDM)

tamr
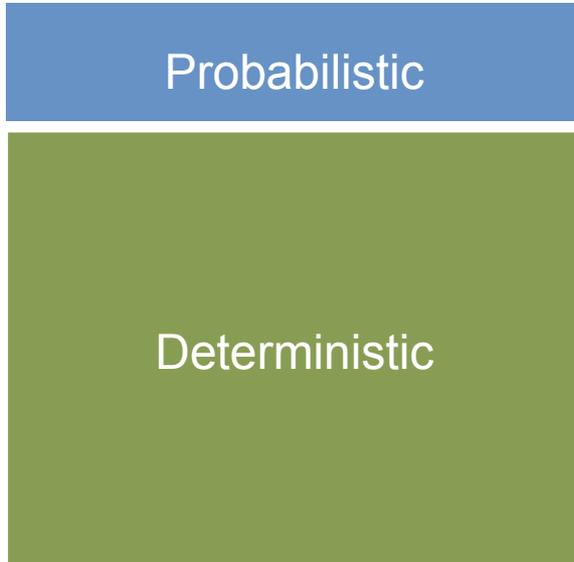
Option #3: "Embrace" Variety -- Tamr's NextGen Approach

Managing enterprise information as an asset requires a new, bottom-up design pattern
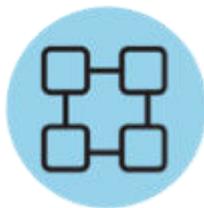
Combine

ALL your metadata and map it to logical entities

Consolidate

Entities and attributes to remove information silos

Classify

Organize your data into an analytics-ready hierarchy

The Two Second Rule.

"Anything that takes a human longer than two seconds is probably unlikely for ML to automatically learn."

- Andrew Ng, Chief Scientist, Baidu