THE TAMR OPPORTUNITY

Explosion of internal data sources, data “lakes” (e.g., Hadoop), external public data sources, has created a tsunami of diverse data sources for enterprises

- Estimated 90% of big data is “dark” data - not used for analytics
- Data scientists spend 80% of time preparing data, leaving 20% for analysis
- Analysts consume as little as 10-12% of the data available to them

Leverage All Data
INTRODUCING TAMR

- Founded in 2013 by enterprise database software veterans
- World-class engineering team
- Top tier venture backing (Google Ventures, NEA)

Jerry Held, PhD
Andy Palmer
Mike Stonebraker, PhD
Ihab Ilyas, PhD
Kevin Willis
Kevin Burke
Nidhi Aggarwal, PhD
Min Xiao
Nik Bates-Haus
RULES BASED APPROACHES DO NOT SCALE WITH DATA VOLUME AND VARIETY

**Traditional rules based Approach**
- Human defines a global schema up front
- Assign a programmer to each data source to understand, map and clean it
- Run rules to transform and de-duplicate the data

**Problems**
- Increasing variety and velocity of data => rules do not work
- No one knows everything about the data
- No automation

**Consequences**
- Data sources are expensive to connect
- <10% of Enterprise data is used for analysis
THIS PROBLEM EXTENDS FAR BEYOND INTERNAL SOURCES

ClinicalTrials.gov is a registry and results database of clinical studies of human participants conducted and funded by clinical studies and about this site, including clinical trials by the U.S. National Institutes of Health.

ClinicalTrials.gov currently lists **173,308 studies** with locations in all 50 states and in **167 countries**.

### Search for Studies

**Example:** "Heart attack" AND "Los Angeles"

- Advanced Search
- See Studies by Topic
- See Studies on a Map

### Search Help

- How to search
- How to find results of studies
- How to read a study record

<table>
<thead>
<tr>
<th>Rank</th>
<th>Status</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Recruiting</td>
<td><strong>Personalized Cancer Therapy</strong>&lt;br&gt;Conditions: Invasive Cancer; Sarcoma; Hematologic Cancer</td>
</tr>
<tr>
<td>2</td>
<td>Terminated</td>
<td><strong>Breast Magnetic Resonance Imaging (MRI) Screening After Mantle Radiation Therapy</strong>&lt;br&gt;Condition: Breast Cancer&lt;br&gt;Interventions: Procedure: Breast MRI; Procedure: Mammogram</td>
</tr>
<tr>
<td>3</td>
<td>Terminated</td>
<td><strong>Anderson Circulating Tumor Cell Burden (CTCB) Study</strong>&lt;br&gt;Conditions: Cancer; Spinal Disease; Vertebral Compression Fractures&lt;br&gt;Interventions: Procedure: Kyphoplasty; Procedure: Vertebroplasty; Device: Cavit</td>
</tr>
<tr>
<td>4</td>
<td>Terminated</td>
<td><strong>TEST: Registry for Endoscopic Head and Neck Surgery</strong>&lt;br&gt;Condition: Head and Neck Cancer&lt;br&gt;Intervention:</td>
</tr>
</tbody>
</table>
The problem is ..... Variety

<table>
<thead>
<tr>
<th>Institution</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>M D Anderson Cancer Center</td>
<td>Diane C. Bodurka</td>
</tr>
<tr>
<td>M D Anderson Cancer Center</td>
<td>Edmund S. Kopetz</td>
</tr>
<tr>
<td>M D Anderson Cancer Center</td>
<td>Elise D. Cook</td>
</tr>
<tr>
<td>University of Texas M.D. Anderson Cancer Center</td>
<td>Elizabeth A. Mittendorf...</td>
</tr>
<tr>
<td>University of Texas M.D. Anderson Cancer Center</td>
<td>Elizabeth A. Mittendorf...</td>
</tr>
<tr>
<td>M.D. Anderson Cancer Center</td>
<td>Elizabeth Shpall, MD</td>
</tr>
<tr>
<td>M D Anderson Cancer Center</td>
<td>Farhad Ravandi-Kashani</td>
</tr>
<tr>
<td>UT M.D. Anderson Cancer Center</td>
<td>Filip Janku, MD</td>
</tr>
<tr>
<td>University of Texas M.D. Anderson Cancer Center</td>
<td>Filip Janku, MD, PhD</td>
</tr>
<tr>
<td>M D Anderson Cancer Center</td>
<td>Gabriel N. Hortobagyi</td>
</tr>
<tr>
<td>The University of Texas M.D. Anderson Cancer Center</td>
<td>Gautam Borthakur, MD</td>
</tr>
<tr>
<td>M D Anderson Cancer Center</td>
<td>George R. Blumenschein</td>
</tr>
</tbody>
</table>
And more variety....

<table>
<thead>
<tr>
<th>Investigator Name</th>
<th>Facility Country</th>
<th>Facility Name</th>
<th>Facility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anjali S. Advani</td>
<td>United States</td>
<td>Hillcrest Hospital Cancer Center</td>
<td>Mayfield</td>
</tr>
<tr>
<td>Anjali S. Advani</td>
<td>United States</td>
<td>Hillcrest Hospital Cancer Center</td>
<td>Mayfield</td>
</tr>
<tr>
<td>Anjali S. Advani</td>
<td>United States</td>
<td>North Coast Cancer Care</td>
<td>Sandusky</td>
</tr>
<tr>
<td><strong>Anjali Advani, MD</strong></td>
<td>United States</td>
<td>Cleveland Clinic</td>
<td>Cleveland</td>
</tr>
<tr>
<td>Anjali S. Advani</td>
<td>United States</td>
<td>Cleveland Clinic Cancer Center - Stron...</td>
<td>Strongly</td>
</tr>
<tr>
<td><strong>Anjali Advani</strong></td>
<td>United States</td>
<td>Cleveland Clinic - Taussig Cancer In...</td>
<td>Cleveland</td>
</tr>
<tr>
<td>Anjali S. Advani</td>
<td>United States</td>
<td>Cleveland Clinic Cancer Center/Fairv...</td>
<td>Cleveland</td>
</tr>
<tr>
<td>Anjali S. Advani</td>
<td>United States</td>
<td>Cleveland Clinic Cancer Center/Fairv...</td>
<td>Cleveland</td>
</tr>
</tbody>
</table>
TAMR EMBRACES DATA VARIETY USING A PROBABILISTIC APPROACH

Catalog
ALL your metadata and map it to logical entities

Connect
Entities and attributes to remove information silos

Curate
Unified data in the application of your choice via APIs
Clinical Trial Data Conversion
A simple, scalable way to automatically convert, validate, and package clinical study data according to the latest CDISC standards.

Procurement Optimization
Fuel procurement optimization through a simplified, unified view of part, supplier, and transaction data.

Customer Data Integration
Drive business growth through a consolidated, 360-degree view of internal and external customer information.

Tamr Vertical Solutions
Procurement Optimization
Fuel optimization through a simple unified view

The Problem
• Procurement managers need complete, accurate, and actionable information to identify cost saving opportunities across the supplier base

• Heterogeneous data created from large investments in enterprise information systems and the lack of referenceable standards makes analysis of long-tail spend opportunities exceedingly difficult

Tamr Solution
• Get a single unified view of your parts, supplier and transaction data

• Let machine learning automate 90% of data matching tasks

• Leverage data expertise in your company to guide data matching
With Solutions For:
Procurement Optimization - Clinical Data Conversion - Customer Data Integration
Thank You

Shobhit Chugh
Product Management and Marketing
shobhit@tamr.com