CAMBRIDGE SEMANTICS
THE SMART DATA COMPANY®

Accelerating Insight

Smart Data Lake Customer Success Stories

Peter Horowitz
Principal - Advisory
PwC

Alok Prasad, President
Ben Szekely, VP – Solution Engineering
Cambridge Semantics

©2017 Cambridge Semantics Inc. All rights reserved. Company Confidential

ACCELERATING INSIGHT

©2017 Cambridge Semantics Inc. All rights reserved.

Anzo Smart Data Lake: Accelerating Insight

A Graph-based Platform to Solve Today's Data and Analytics Challenges

Anzo Smart Data Lake Value Proposition

All Data Sources

Enterprise Data Lakes

Data Ingestion & Mapping

Load and Map

any data…….
from anywhere…
at any time!

Data Cataloging

Organize, Integrate, Share and Govern

any data assets..
more efficiently
greater security

In-Memory MPP Query

Graph Exploration

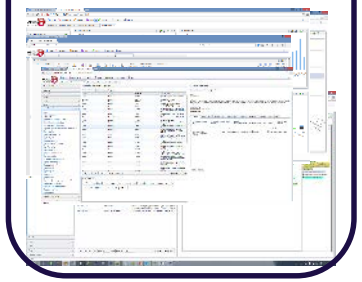
Interconnected …
trillions of facts…
100x the speed…

Discovery & Analytics

Ask & Answer

any questions…
any combination.
any user…..

Actionable Insights



"Last Mile" Analytics

tableau
TIBCO Spotfire
R
sas

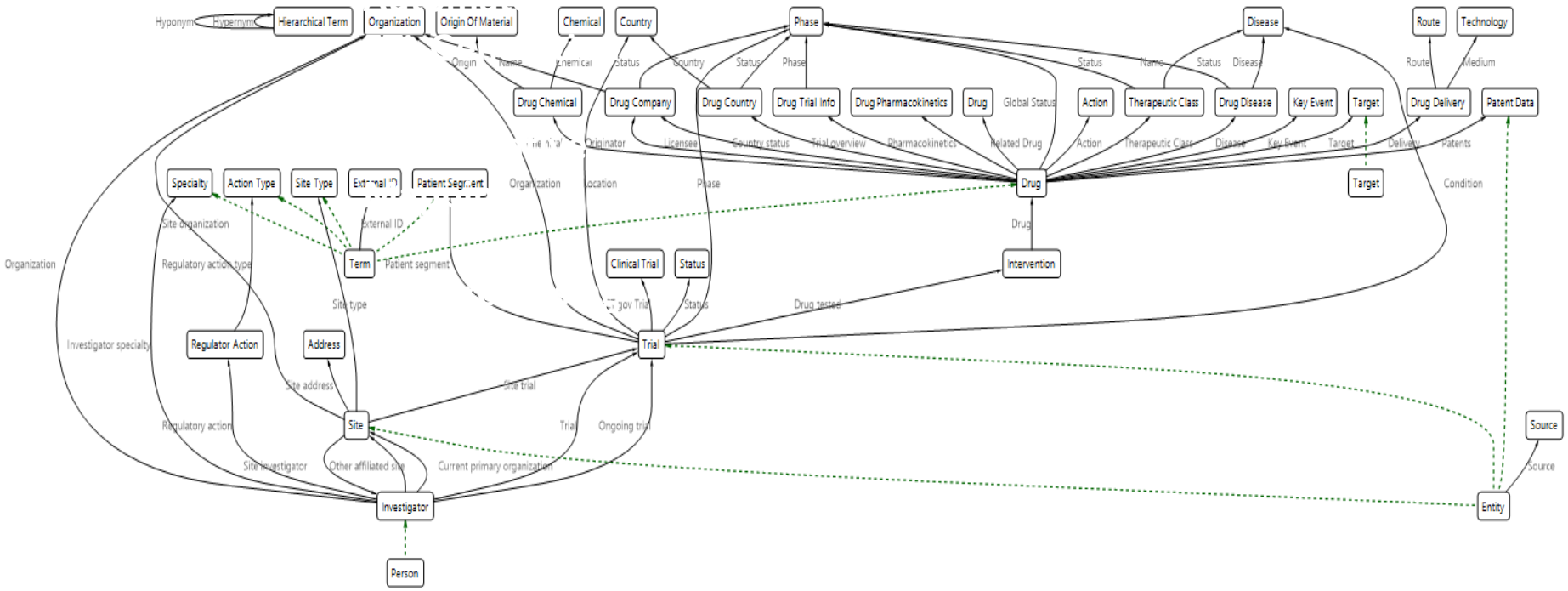
Allowing Enterprises to Rapidly Create Purpose-built Discovery & Analytics Solutions on the Cloud of Choice

©2017 Cambridge Semantics Inc. All rights reserved.

Anzo Smart Data Lake: Accelerating Insight

A Graph-based Platform to Solve Today's Data and Analytics Challenges

Enterprise Knowledge Graph



Accelerated Insights – 100x faster than the nearest competitor

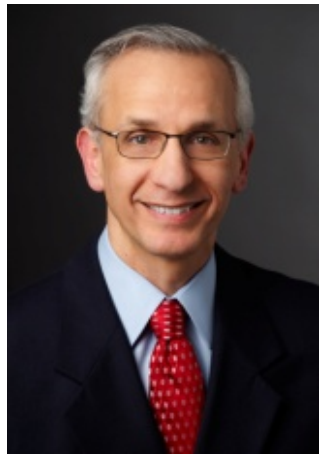
©2017 Cambridge Semantics Inc. All rights reserved.

Sample Customers

©2017 Cambridge Semantics Inc. All rights reserved.

Accelerating Insight – Customer Successs Story

Peter Horowitz
Principal – Advisory Banking and Capital Management
PwC



Ben Szekely
VP Solution Engineering
Cambridge Semantics

©2017 Cambridge Semantics Inc. All rights reserved.

Accelerating Insights: Using Smart Data Lakes Successfully

March 2017

pwc

CAMBRIDGE SEMANTICS
THE SMART DATA COMPANY®

Overview

The Dilemma *– Given current landscape within many financial service institutions, the standard data journey takes* ***TOO LONG***

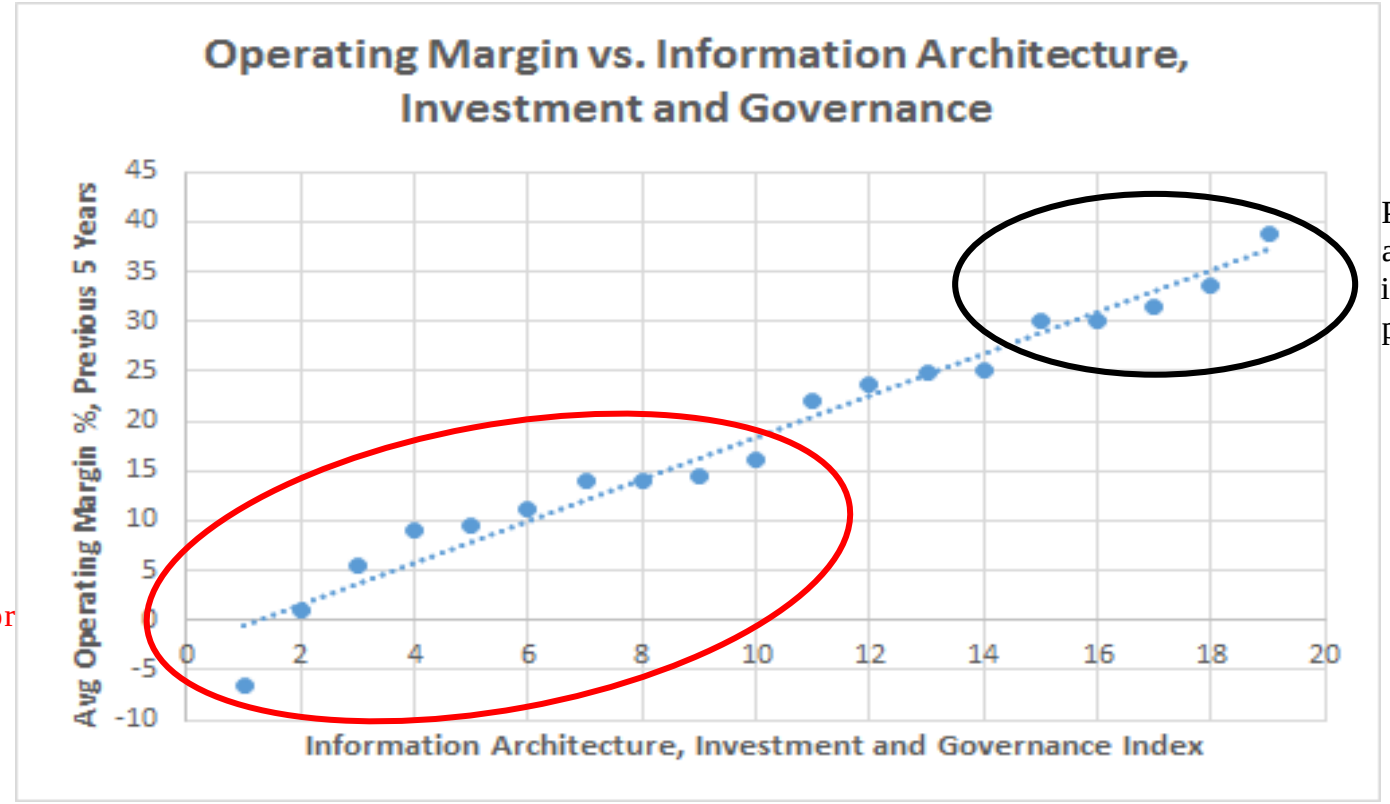
Winning Strategy *- Attack every step in the process to reduce* ***time to market.***

Typical data journey

Why is this so important?

For global top banks, comparing average operating margin percent for previous 5 years against index of information architecture, investment and governance demonstrates strong correlation.

Proper investment in information strategy, architecture and governance *PAYS OFF*!



Proper data architecture improving or in place.

Insufficient or lagging data architecture investment, execution, governance.

Sources: Average OM% from Morningstar, Inc.; IAIG Index

Old approach is Challenged

The existing landscape presents great challenges since it is based on relational databases and point-to-point mappings.

Challenges

Data source mapping

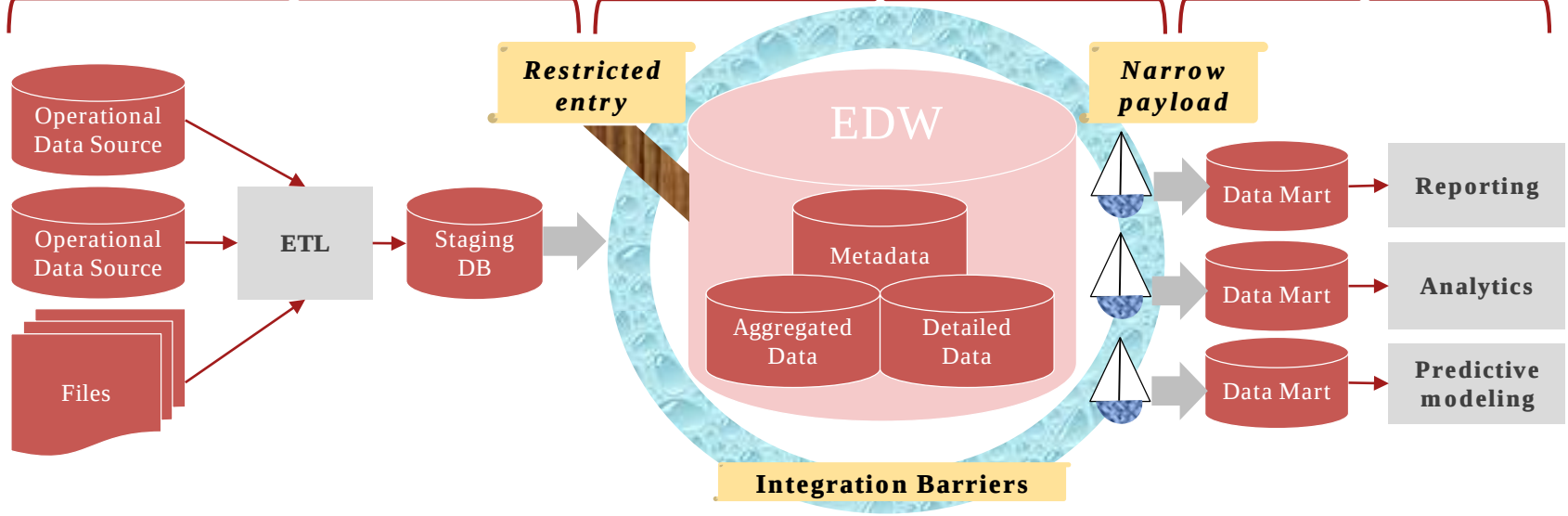
Multi-stage, brittle ETL required to integrate diverse data sources

Standard model required

Enforces a standard model, no exceptions, leading to loss of valuable information context

Limited insights

Rigidity undermines the agility required in data analytics



New approach

The smart data lake provides a centralized data store with a flexible schema where data is first loaded and then transformed; a process known as schema-on-read.

Benefits

Original data and models maintained

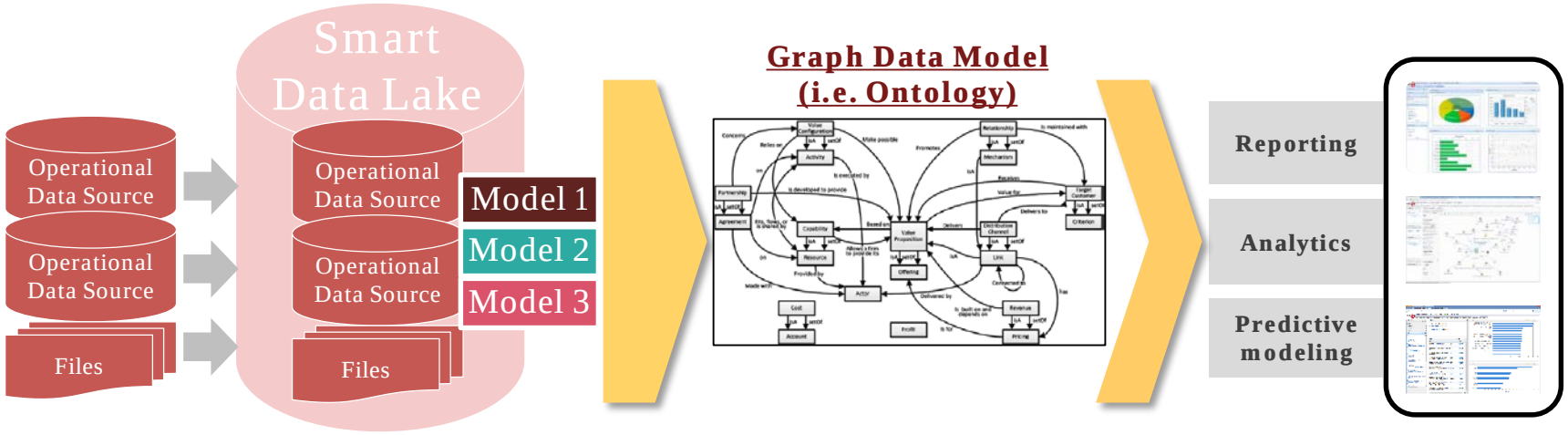
Data is loaded seamlessly without transformation or mapping, irrespective of format

Unified model

Consolidated model unifies disparate source models into a comprehensive and shared canonical model

Richer insights

Cross region and domain view supports population-based analytics, hidden relationship discovery and reduces risk of siloed or inaccessible



Graph store designs capture the nature of data relationships and supports a variety of data representations, known as "Shared Semantics"

Case Study: *Next generation insider trading and fraud surveillance driven by structured and unstructured data*

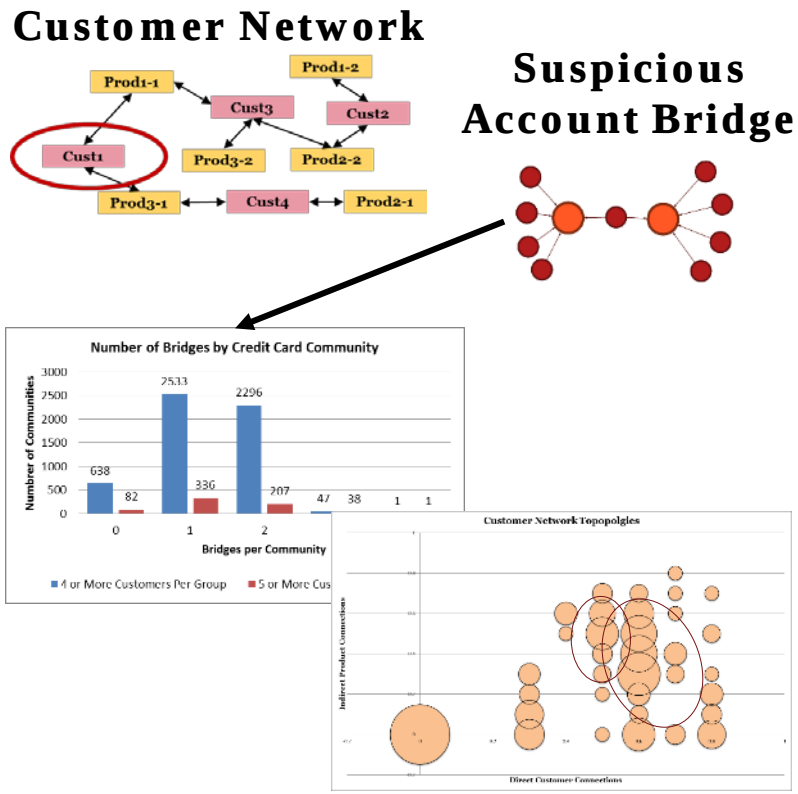
Analysis Overview

- Ingested emails, IMs, web browser logs, cookies, phone records, contacts, news feeds
- Ingested futures, options trades, market price data

Impact

- Data was ingested and linked in hours.
- Profiles developed for employee roles
- Outlier behavior effortlessly identified
- Employees given risk ratings allowing for prioritization of detailed compliance investigation.

Cast Study: *Terabytes of data ingested and analyzed in a fraction of the time revealing complex relationships between clients and transactions*



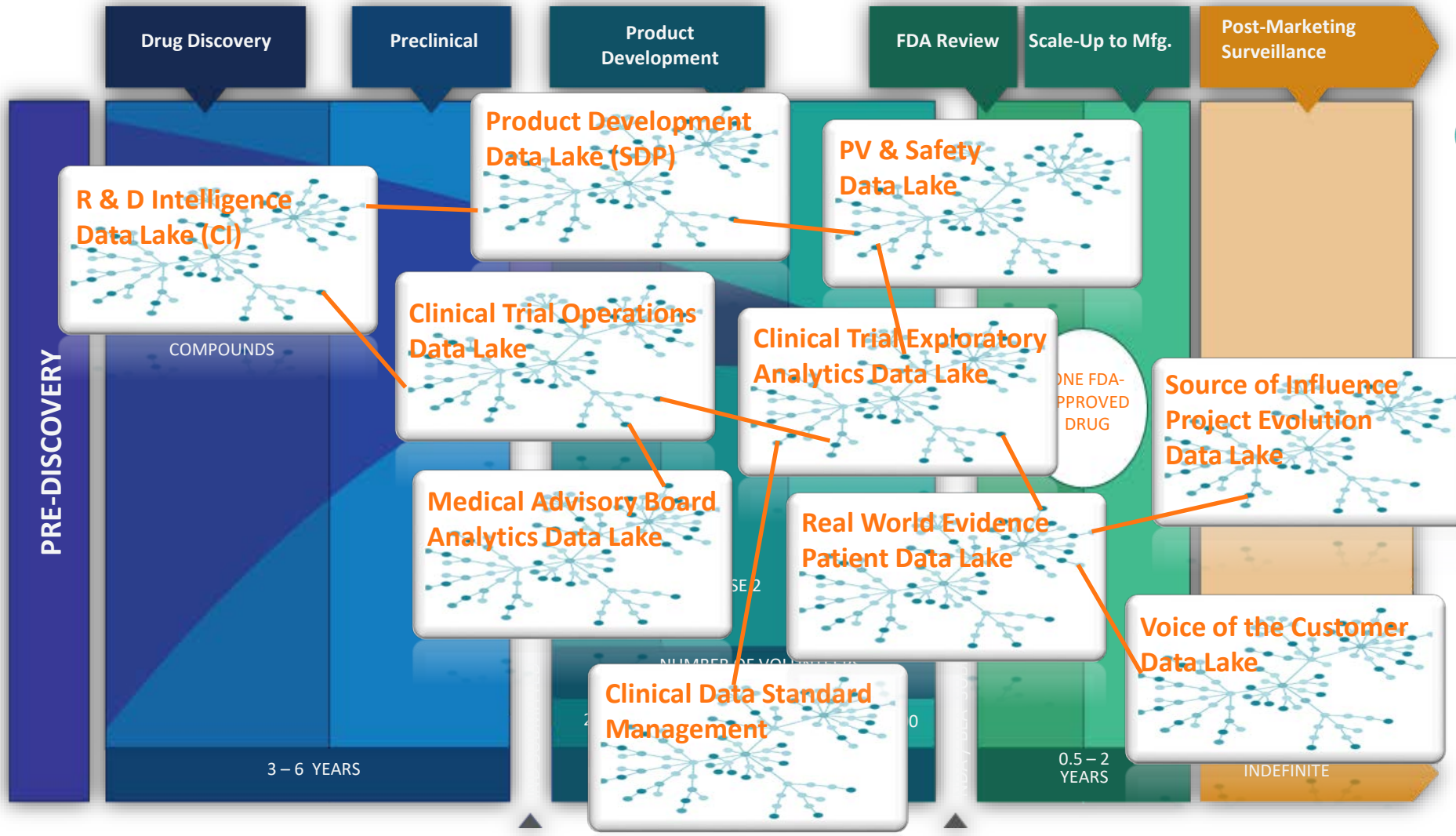
Analysis Overview

- 5.5 billion transactions on behalf of 25 million customers for 35 million accounts.
- Based on reference data and public information, prepared “social” network of clients analyzing edges (i.e., connections) between them.

Impact

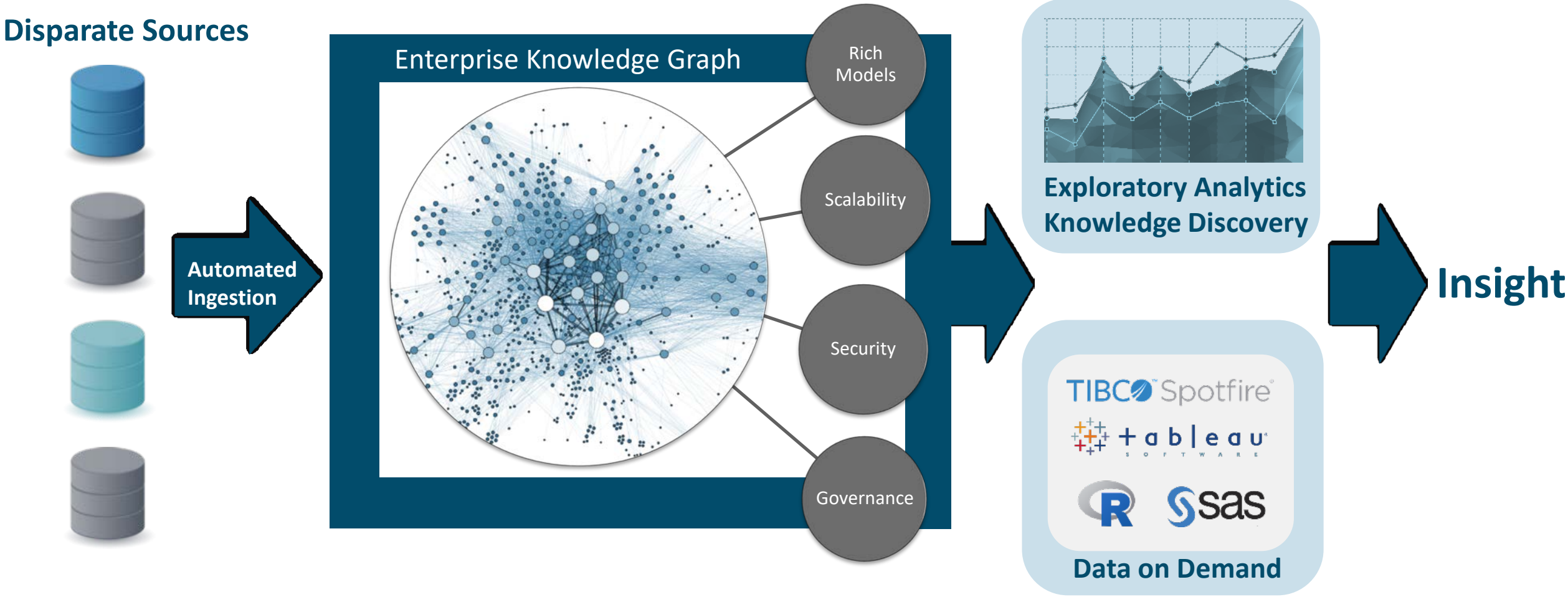
- Data was ingested and linked in hours.
- Edges developed and weighted based on Pointwise Mutual Information algorithms.
- Resulting graph database supported recursive analysis, which would have been dramatically longer using RDBS, allowing for identification of suspicious networks

Anzo Smart Data Lakes Across Pharma



©2017 Cambridge Semantics Inc. All rights reserved.

Anzo Smart Data Lake: Accelerating Insight



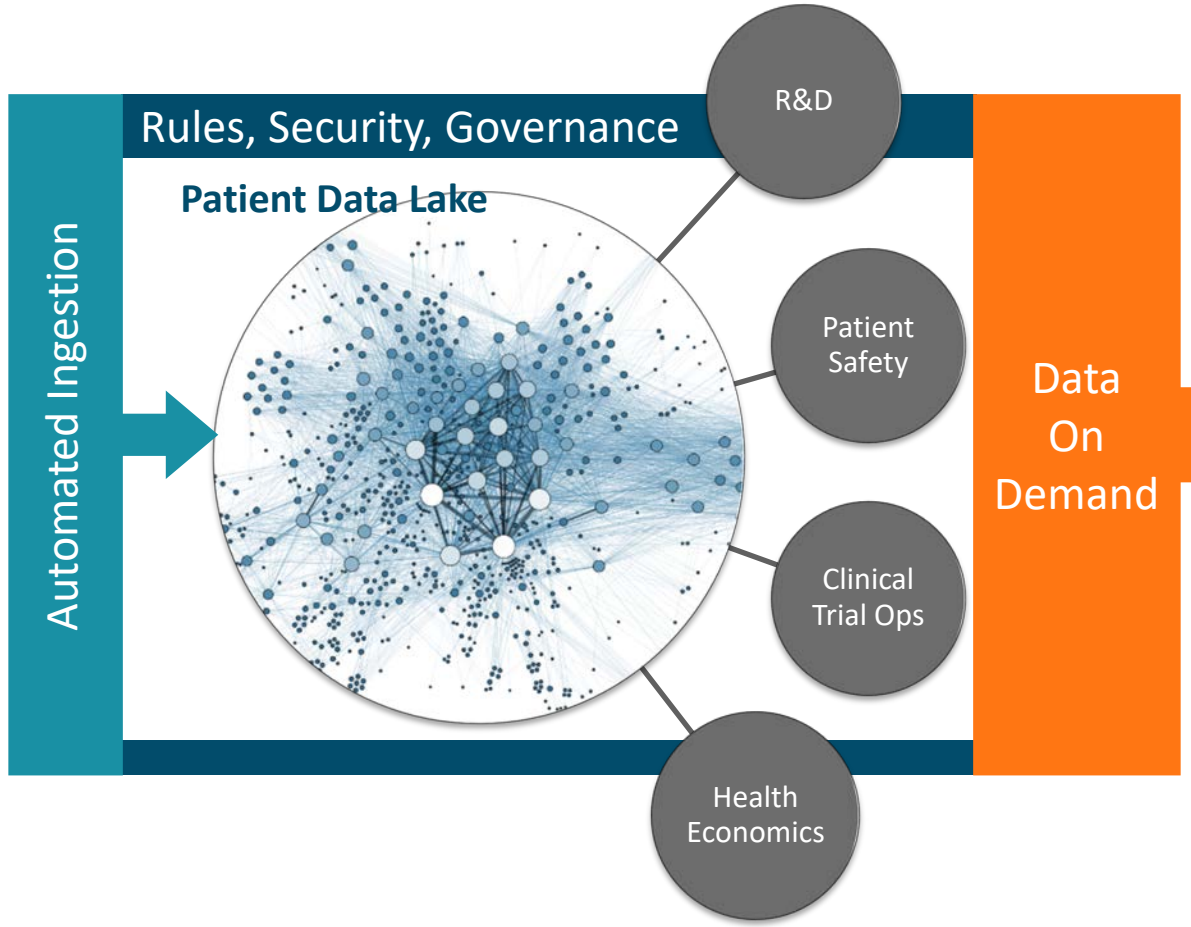
©2017 Cambridge Semantics Inc. All rights reserved.

The Smart Patient Data Lake

Connectors → Models → Rules → Analytics & Tools

Patient Data Systems

- Insurance Claims
- Clinical Trials
- Rx Data
- Health Records
- Genetic Data



Data On Demand

Insight for Decision Makers

Improving patient outcomes, safety,and comfort

Reducing the time bring medicines to patients

Lowering the cost of healthcare

©2017 Cambridge Semantics Inc. All rights reserved.

Manage Working Set +

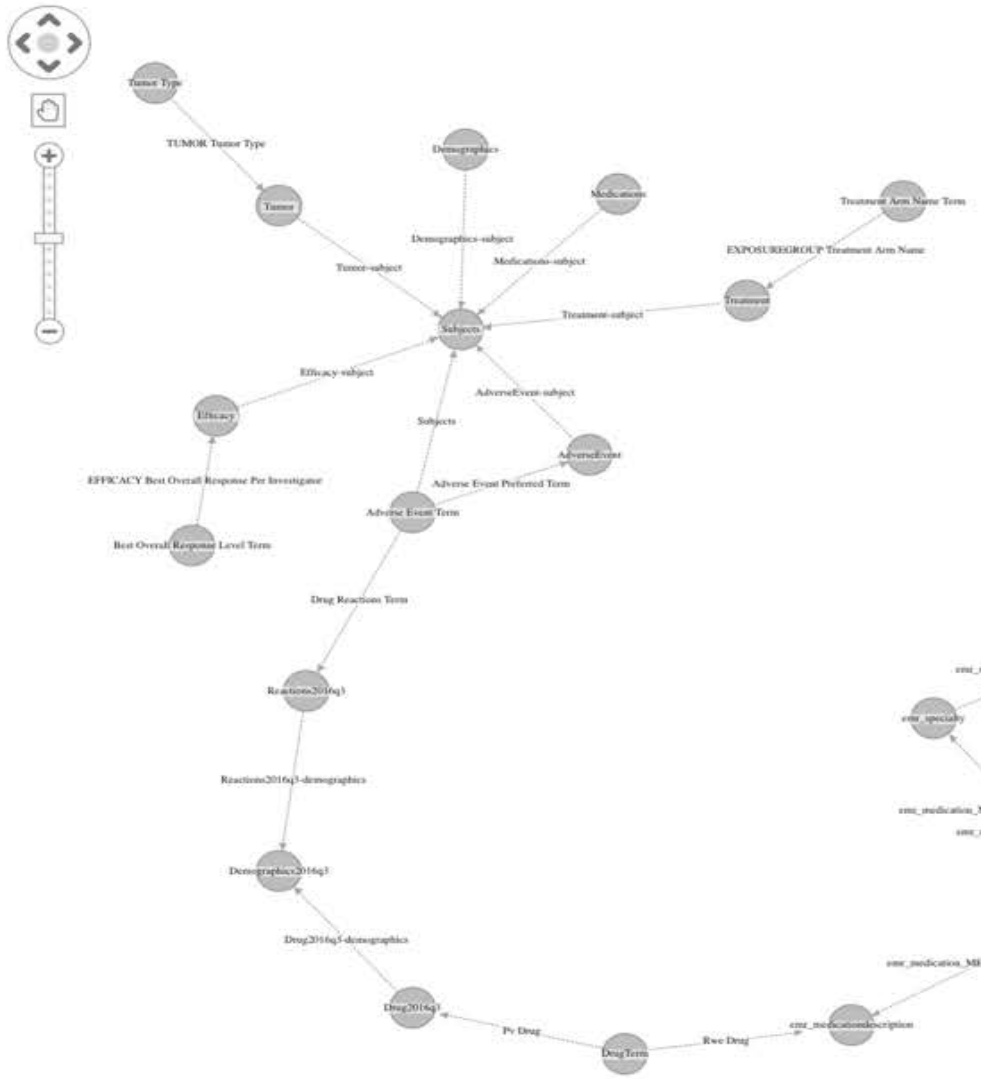
Filter Ontology Items

- emr_complaintdescription
  - COMPLAINTID
  - COMPLAINTSTRING
- emr_complaintdescriptio...
- emr_medication
  - emr_medication_PAT...
  - emr_medication_ME...
  - emr_medication_ME...
  - emr_medication_ME...
  - MEDICATIONSPECI...
  - MEDICATIONDRUG...
  - emr_medication_ME...
  - emr_medication_ME...
  - MEDICATIONACTIV...
  - MEDICATIONACTIV...
  - MEDICATIONCREAT...
  - MEDICATIONAGE
  - MEDICATIONSIGNA...
  - MEDICATIONSTART...
  - MEDICATIONDRUG...
  - MEDICATIONSTUD...
  - PATIENTID

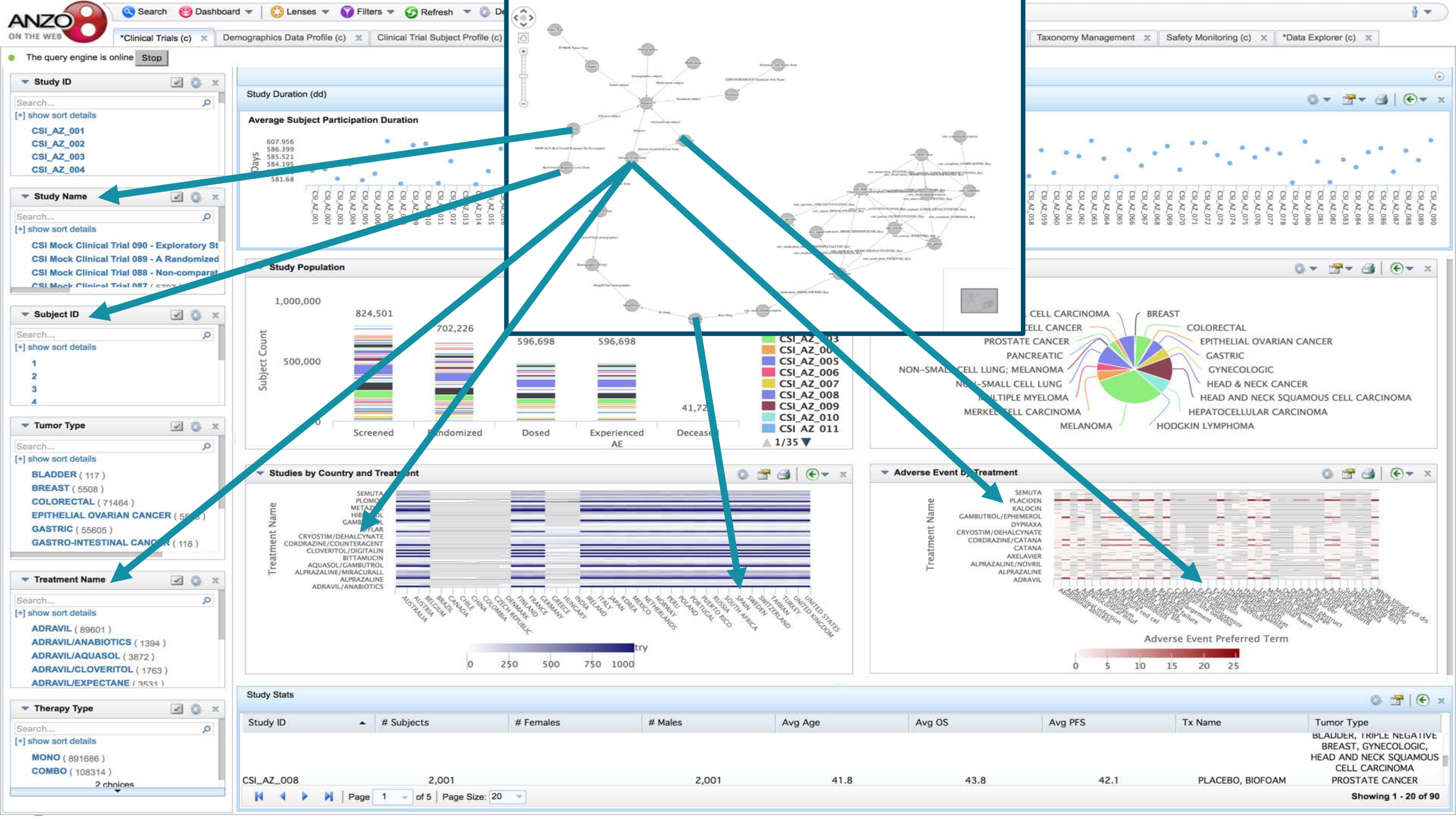
CREATE ONTOLOGY

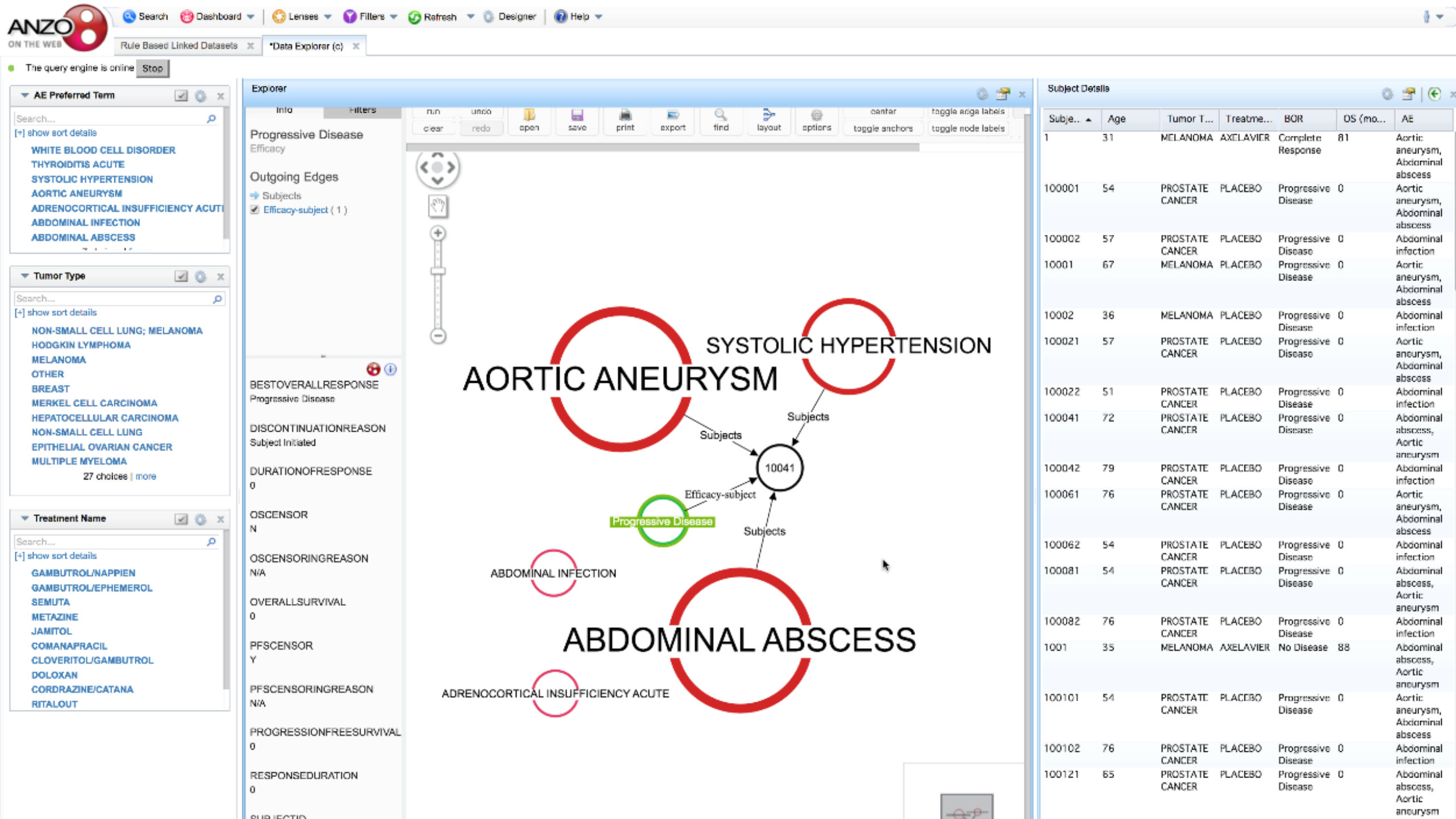
Edit

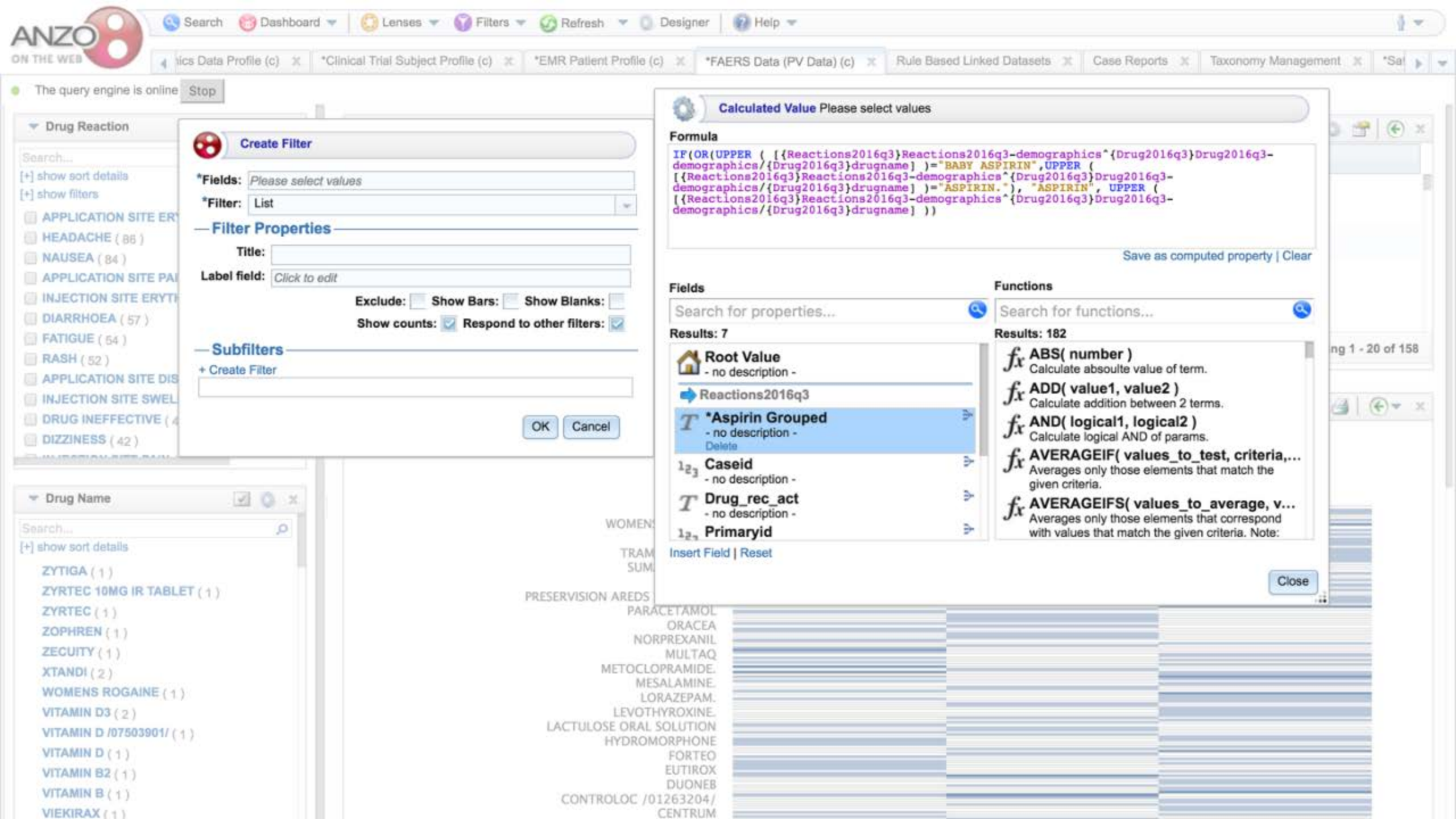
View



SAVE CANCEL


ANZO ON THE WEB
Search
Dashboard
Lenses
Filters
Refresh
*Clinical Trials (c)
Demographics Data Profile (c)
Clinical Trial Subject Profile (c)
Taxonomy Management
Safety Monitoring (c)
*Data Explorer (c)
The query engine is online
Stop
Study ID
Search...
[+] show sort details
CSI_AZ_001
CSI_AZ_002
CSI_AZ_003
CSI_AZ_004
Study Name
CSI Mock Clinical Trial 090 - Exploratory St
CSI Mock Clinical Trial 089 - A Randomized
CSI Mock Clinical Trial 088 - Non-comparat
Subject ID
1
2
3
4
Tumor Type
BLADDER (117)
BREAST (5508)
COLORECTAL (71464)
EPITHELIAL OVARIAN CANCER
GASTRIC (55605)
Treatment Name
ADRAVIL (89601)
ADRAVIL/ANABIOTICS (1394)
ADRAVIL/AQUASOL (3872)
ADRAVIL/CLOVERITOL (1763)
ADRAVIL/EXPECTANE
Therapy Type
MONO (891686)
COMBO (108314)
2 choices
Study Duration (dd)
Average Subject Participation Duration
Days
607.956
586.399
585.521
584.195
581.68
Study Population
Subject Count
1,000,000
500,000
824,501
702,226
596,698
596,698
41,72
Screened
Dosed
Experienced AE
CSI_AZ_005
CSI_AZ_006
CSI_AZ_007
CSI_AZ_008
CSI_AZ_009
CSI_AZ_010
CSI_AZ_011
1/35
BREAST
COLORECTAL
EPITHELIAL OVARIAN CANCER
GASTRIC
GYNECOLOGIC
HEAD & NECK CANCER
HEAD AND NECK SQUAMOUS CELL CARCINOMA
HEPATOCELLULAR CARCINOMA
HODGKIN LYMPHOMA
MELANOMA
PROSTATE CANCER
PANCREATIC
Studies by Country and Treatment
Treatment Name
SEMUTA
CRYOSTIM/DEHALCYNATE
CORDRAZINE/COUNTERAGENT
CLOVERITOL/DIGITALIN
BITTAMUCIN
AQUASOL/GAMBUTROL
ALPRAZALINE/MIRACURALL
ALPRAZALINE
ADRAVIL/ANABIOTICS
AUSTRALIA
BELGIUM
BRAZIL
CANADA
CHILE
CHINA
COLOMBIA
CZECH REPUBLIC
DENMARK
FINLAND
FRANCE
GERMANY
GREECE
HUNGARY
INDIA
IRELAND
ITALY
JAPAN
KOREA
MEXICO
NETHERLANDS
NORWAY
PERU
POLAND
PORTUGAL
PUERTO RICO
RUSSIA
SOUTH AFRICA
SPAIN
SWEDEN
SWITZERLAND
TAIWAN
TURKEY
UNITED KINGDOM
0
250
500
750
1000
Adverse Event by Treatment
Treatment Name
SEMUTA
PLACIDEN
KALOCIN
GAMBUTROL/EPHEMEROL
DYPRAXA
CRYOSTIM/DEHALCYNATE
CORDRAZINE/CATANA
CATANA
AXELAVIER
ALPRAZALINE/NOVRIL
ALPRAZALINE
ADRAVIL
Adverse Event Preferred Term
0
5
10
15
20
25
Study Stats
Study ID
Subjects
Females
Males
Avg Age
Avg OS
Avg PFS
Tx Name
Tumor Type
CSI_AZ_008
2,001
2,001
41.8
43.8
42.1
PLACEBO, BIOFOAM
BLADDER, TRIPLE NEGATIVE BREAST, GYNECOLOGIC, HEAD AND NECK SQUAMOUS CELL CARCINOMA PROSTATE CANCER
Page 1 of 5
Page Size: 20
Showing 1 - 20 of 90


ANZO 8 ON THE WEB
Search
Dashboard
Lenses
Filters
Refresh
Designer
Help
Rule Based Linked Datasets
*Data Explorer (c)
The query engine is online
Stop
AE Preferred Term
Search...
[+] show sort details
WHITE BLOOD CELL DISORDER
THYROIDITIS ACUTE
SYSTOLIC HYPERTENSION
AORTIC ANEURYSM
ADRENOCORTICAL INSUFFICIENCY ACUTE
ABDOMINAL INFECTION
ABDOMINAL ABSCESS
Tumor Type
Search...
[+] show sort details
NON-SMALL CELL LUNG; MELANOMA
HODGKIN LYMPHOMA
MELANOMA
OTHER
BREAST
MERKEL CELL CARCINOMA
HEPATOCELLULAR CARCINOMA
NON-SMALL CELL LUNG
EPITHELIAL OVARIAN CANCER
MULTIPLE MYELOMA
27 choices | more
Treatment Name
Search...
[+] show sort details
GAMBUTROL/NAPPIEN
GAMBUTROL/EPHEMEROL
SEMUTA
METAZINE
JAMITOL
COMANAPRACIL
CLOVERITOL/GAMBUTROL
DOLOXAN
CORDRAZINE/CATANA
RITALOUT
Explorer
Info
Filters
Progressive Disease
Efficacy
Outgoing Edges
Subjects
Efficacy-subject (1)
BESTOVERALLRESPONSE
Progressive Disease
DISCONTINUATIONREASON
Subject Initiated
DURATIONOFRESPONSE
0
OSCENSOR
N
OSCENSORINGREASON
N/A
OVERALLSURVIVAL
0
PFSCENSOR
Y
PFSCENSORINGREASON
N/A
PROGRESSIONFREESURVIVAL
0
RESPONSEDURATION
0
run
undo
clear
redo
open
save
print
export
find
layout
options
center
toggle anchors
toggle edge labels
toggle node labels
SYSTOLIC HYPERTENSION
AORTIC ANEURYSM
Subjects
Subjects
10041
Efficacy-subject
Progressive Disease
Subjects
ABDOMINAL INFECTION
ABDOMINAL ABSCESS
ADRENOCORTICAL INSUFFICIENCY ACUTE
Subject Details
Subje... Age Tumor T... Treatme... BOR OS (mo... AE
1 31 MELANOMA AXELAVIER Complete Response 81 Aortic aneurysm, Abdominal abscess
100001 54 PROSTATE CANCER PLACEBO Progressive Disease 0 Aortic aneurysm, Abdominal abscess
100002 57 PROSTATE CANCER PLACEBO Progressive Disease 0 Abdominal infection
10001 67 MELANOMA PLACEBO Progressive Disease 0 Aortic aneurysm, Abdominal abscess
10002 36 MELANOMA PLACEBO Progressive Disease 0 Abdominal infection
100021 57 PROSTATE CANCER PLACEBO Progressive Disease 0 Aortic aneurysm, Abdominal abscess
100022 51 PROSTATE CANCER PLACEBO Progressive Disease 0 Abdominal infection
100041 72 PROSTATE CANCER PLACEBO Progressive Disease 0 Abdominal abscess, Aortic aneurysm
100042 79 PROSTATE CANCER PLACEBO Progressive Disease 0 Abdominal infection
100061 76 PROSTATE CANCER PLACEBO Progressive Disease 0 Aortic aneurysm, Abdominal abscess
100062 54 PROSTATE CANCER PLACEBO Progressive Disease 0 Abdominal infection
100081 54 PROSTATE CANCER PLACEBO Progressive Disease 0 Abdominal abscess, Aortic aneurysm
100082 76 PROSTATE CANCER PLACEBO Progressive Disease 0 Abdominal infection
1001 35 MELANOMA AXELAVIER No Disease 88 Abdominal abscess, Aortic aneurysm
100101 54 PROSTATE CANCER PLACEBO Progressive Disease 0 Aortic aneurysm, Abdominal abscess
100102 76 PROSTATE CANCER PLACEBO Progressive Disease 0 Abdominal infection
100121 65 PROSTATE CANCER PLACEBO Progressive Disease 0 Abdominal abscess, Aortic aneurysm


ANZO ON THE WEB
Search
Dashboard
Lenses
Filters
Refresh
Designer
Help
*FAERS Data (PV Data) (c)
Rule Based Linked Datasets
Case Reports
Taxonomy Management
The query engine is online
Stop
Drug Reaction
Create Filter
*Fields:
Please select values
*Filter:
List
Filter Properties
Title:
Label field:
Click to edit
Exclude:
Show Bars:
Show Blanks:
Show counts:
Respond to other filters:
Subfilters
+ Create Filter
OK
Cancel
Calculated Value Please select values
Formula
IF(OR(UPPER ([{Reactions2016q3}Reactions2016q3-demographics^{Drug2016q3}Drug2016q3-demographics/{Drug2016q3}drugname])="BABY ASPIRIN",UPPER ([{Reactions2016q3}Reactions2016q3-demographics^{Drug2016q3}Drug2016q3-demographics/{Drug2016q3}drugname])="ASPIRIN."), "ASPIRIN", UPPER ([{Reactions2016q3}Reactions2016q3-demographics^{Drug2016q3}Drug2016q3-demographics/{Drug2016q3}drugname]))
Save as computed property | Clear
Fields
Search for properties...
Results: 7
Root Value
- no description -
Reactions2016q3
*Aspirin Grouped
- no description -
Delete
Caseid
- no description -
Drug_rec_act
- no description -
Primaryid
Functions
Search for functions...
Results: 182
ABS(number)
Calculate absoulte value of term.
ADD(value1, value2)
Calculate addition between 2 terms.
AND(logical1, logical2)
Calculate logical AND of params.
AVERAGEIF(values_to_test, criteria,...
Averages only those elements that match the given criteria.
AVERAGEIFS(values_to_average, v...
Averages only those elements that correspond with values that match the given criteria. Note:
Insert Field | Reset
Close
Drug Name
ZYTIGA (1)
ZYRTEC 10MG IR TABLET (1)
ZYRTEC (1)
ZOPHREN (1)
ZECUITY (1)
XTANDI (2)
WOMENS ROGAINE (1)
VITAMIN D3 (2)
VITAMIN D /07503901/ (1)
VITAMIN D (1)
VITAMIN B2 (1)
VITAMIN B (1)
VIEKIRAX (1)
PRESERVISION AREDS
PARACETAMOL
ORACEA
NORPREXANIL
MULTAQ
METOCLOPRAMIDE.
MESALAMINE.
LORAZEPAM.
LEVOTHYROXINE.
LACTULOSE ORAL SOLUTION
HYDROMORPHONE
FORTEO
EUTIROX
DUONEB
CONTROLOC /01263204/
CENTRUM

ANZO ON THE WEB

Search | Dashboard | Lenses | Filters | Refresh | Designer | Help

FAERS Data (PV Data) (c) | *Case Reports | Rule Based Linked Datasets | *Data Explorer (c) | Diagnoses and Meds Annotations

Linked Data Sets

Data Types

Document

Search...

[+] show sort details

- Safety Labeling Changes Approved By
- RECALL OF OCTAPERBROONIUM.pdf
- Prempro 2003 Safety Alert.pdf
- NATIONWIDE ALERT ON ZOMBREX IN
- Morpha Safety Update.pdf
- Medwatch Carabonal Letter.pdf
- MedWatch - AQUASOL Letter to Clinic
- MedWatch - ADRAVIL Dear Pharmacis
- Lipitor 2003 Safety Alert.pdf
- Lipitor 2003 Safety Alert 2.pdf

31 choices | more

Annotation

Search...

[+] show sort details

- BLOOD
- RENAL
- ATORVASTATIN
- PLASMA
- AMPHETAMINE
- MEDICATION
- BREAST CANCER
- DIETARY SUPPLEMENT
- ENSURE
- ARTHRITIS
- STEROIDS
- INFECTION
- QUININE
- PROTEIN

Annotator Name

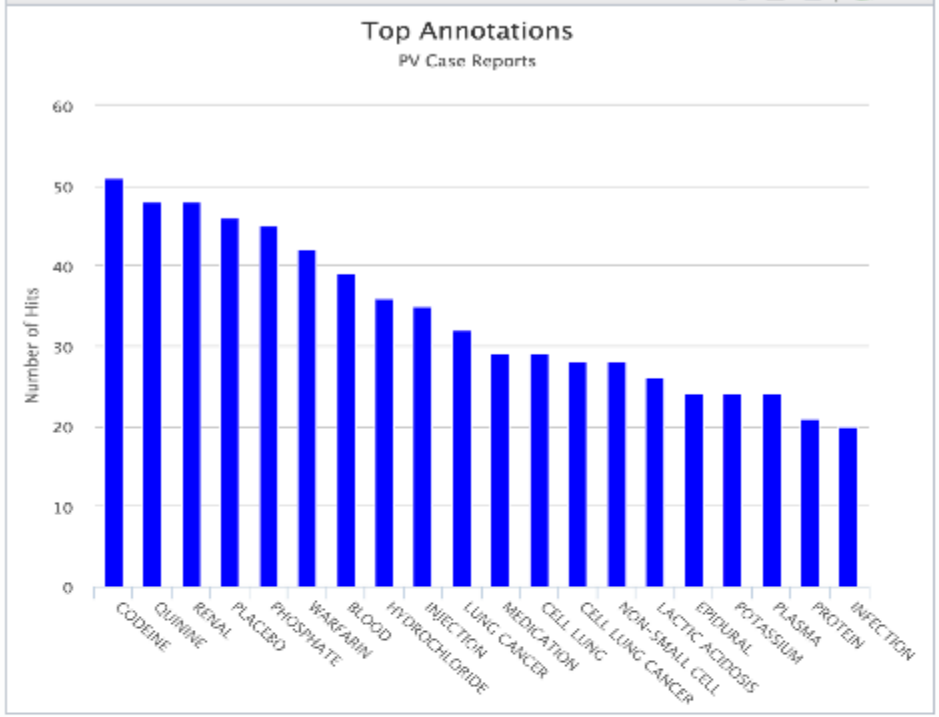
Search...

[+] show sort details

- Significant (380)
- KB (707)

2 choices

Significant Phrases



Extracted Annotations

Document	Category	Instance	Hits	Context
Bittamucin FDA posts safety alerts.pdf	Annotations	WARFARIN	42	...n patients receiving Bittamucin concurrently with warfarin. Since Bittamucin was approved by the FDA on Janu...tamucin did not alter the anticoagul..
Safety Labeling Changes Approved By FDA Center for Drug Evaluation and Research.pdf	Annotations	PHOSPHATE	36	...chloride, phenylephrine hydrochloride and codeine phosphate is contraindicated in pediatric patients less tha...chloride, phenylephrine hydrochlori..
IMPORTANT PRESCRIBING INFORMATION UPSIRIM.pdf	Annotations	QUININE	33	... data indicate that although beverages containing quinine (up to 80 mg/day or about 32 ounces of tonic wate...e of ketoconazole, itraconazole, cryt..

Annotation

Annotated Document

DOCUMENT: Bittamucin FDA posts safety alerts.pdf (Bittamucin FDA posts safety alerts.pdf)

LAST MODIFIED: 2/22/2017 10:15PM

ulation.

atients receiving Bittamucin concurrently with warfarin there have been infre
keting reports of increases in prothrombin time, sometimes associated with b
dominantly in the elderly. As you know, the warfarin label advises close and
nagement of warfarin "whenever other medications are initiated, discontinuec
gularly," given the increased potential for bleeding. The warfarin label indica
ractions with more than a hundred compounds.

amucin is indicated for treatment of NON-HODGKIN LYMPHOMA.

amucin, by itself, has no effect on platelet aggregation or bleeding time at the
be an appropriate treatment option in patients taking warfarin, when cancer
cated. It is important to note that Bittamucin can be continued or initiated wi
ents are monitored appropriately for changes in anticoagulant activity, partic
days. It is estimated that approximately one percent of patients on Bittamuci
farin, with a smaller subset of that group reporting increases in prothrombin t

rder to communicate this important information to healthcare professionals,
ion under the PRECAUTIONS heading in the revised package insert for Bitt
losed for full revised label) now reads:

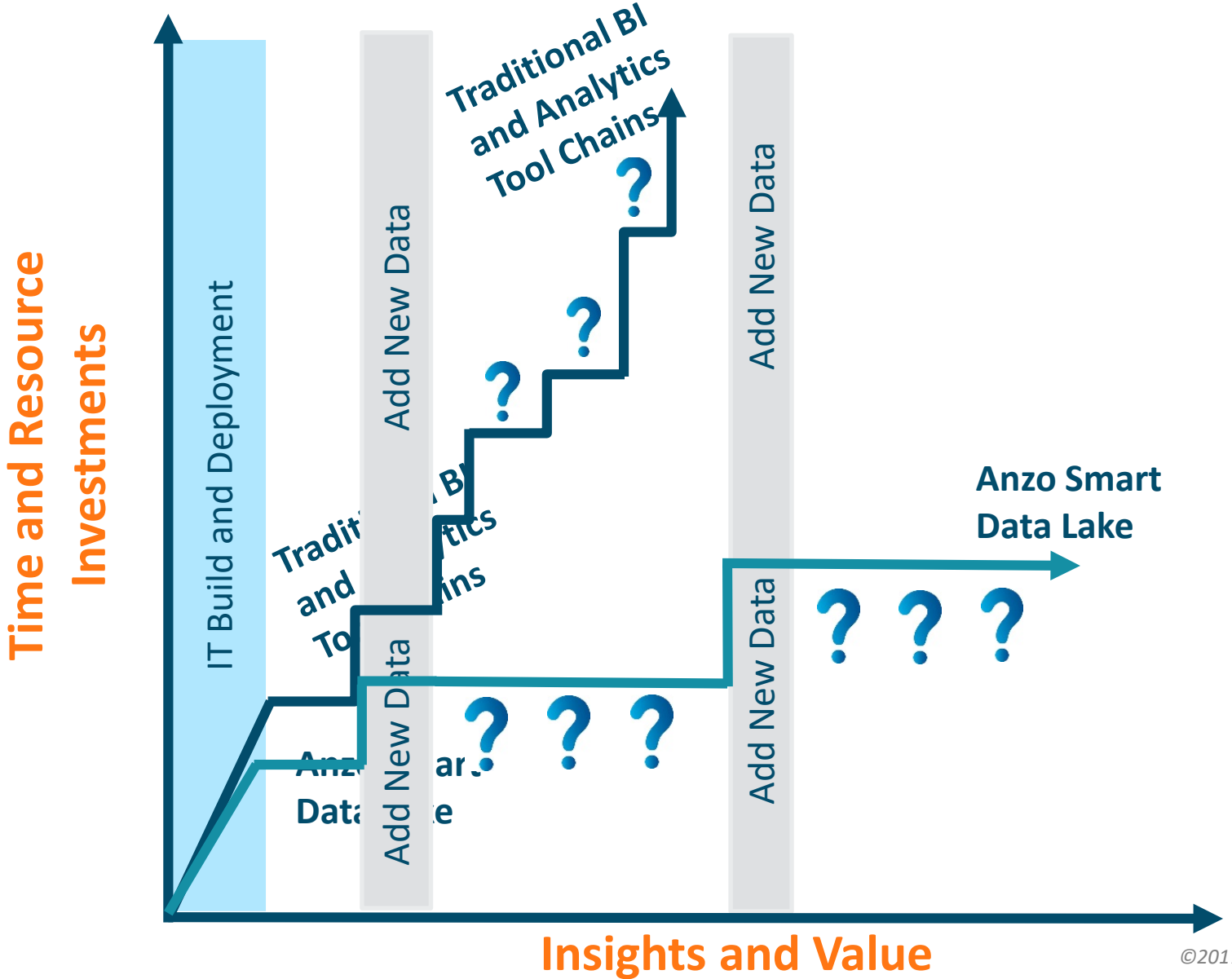
Warfarin: Anticoagulant activity should be monitored, particularly in the
after initiating or changing Bittamucin therapy in patients receiving warfa
agents, since these patients are at an increased risk of bleeding complicati
Bittamucin on the anticoagulant effect of warfarin was studied in a group
receiving daily doses of 2 to 5 mg of warfarin. In these subjects, Bittamuc
anticoagulant effect of warfarin as determined by prothrombin time. Howe
marketing experience, bleeding events have been reported, predominantly
association with increases in prothrombin time in patients receiving Bittar
with warfarin.

ce Bittamucin was approved by the FDA on January 31, 2006, more than 4 m
e been dispensed. The overall general safety profile of Bittamucin remains u

Concept Type Key

- [KB: Annotations] (48)
- [Significant: Significant Phrase] (22)

Disrupting the Time-to-Value Function



©2017 Cambridge Semantics Inc. All rights reserved.

Anzo Smart Data Lake

A Graph-based Platform to Disrupt Today's Data and Analytics Challenges

Enterprise Data Sources

Enterprise Data Lakes

Data Ingestion & Mapping

- Automated ETL Generation
- Collaborative Mapping
- Text Processing

Data Cataloging

- Data & Model Governance
- Active Metadata Management
- Role-Based Security

In-Memory MPP Query

- Graphmarts on Demand
- ELT, Model Based Data Integration
- Document Search

Discovery & Analytics

- Automated Query Generation
- User Dashboards and Custom UI/UX
- Self-Serve Live Extracts

Actionable Insights

"Last Mile" Analytics

tableau

TIBCO Spotfire

R

SAS

Connectors → Models → Rules → Analytics & Tools

ASDL Customer Fingerprint - Intellectual Property

©2017 Cambridge Semantics Inc. All rights reserved.

Anzo Smart Data Lake: Accelerating Insight

A Graph-based Platform to Solve Today's Data and Analytics Challenges

Anzo Smart Data Lake
Select Capabilities

All Data Sources

Enterprise Data Lakes

Data Ingestion & Mapping

- Automated ETL Generation
- Collaborative Mapping
- Text Processing

Data Cataloging

- Data & Model Governance
- Active Metadata Management
- Role-Based Security

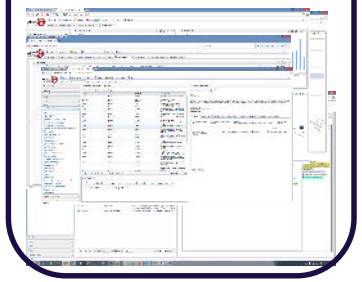
In-Memory MPP Query

- Graphmarts on Demand
- ELT, Model Based Data Integration
- Document Search

Discovery & Analytics

- Knowledge Discovery
- Custom User Dashboards
- Self-Serve Live Extracts

Actionable Insights



"Last Mile" Analytics

tableau
TIBCO Spotfire
R
sas

Allowing Enterprises to Rapidly Create Purpose-built Discovery & Analytics Solutions on the Cloud of Choice

©2017 Cambridge Semantics Inc. All rights reserved.