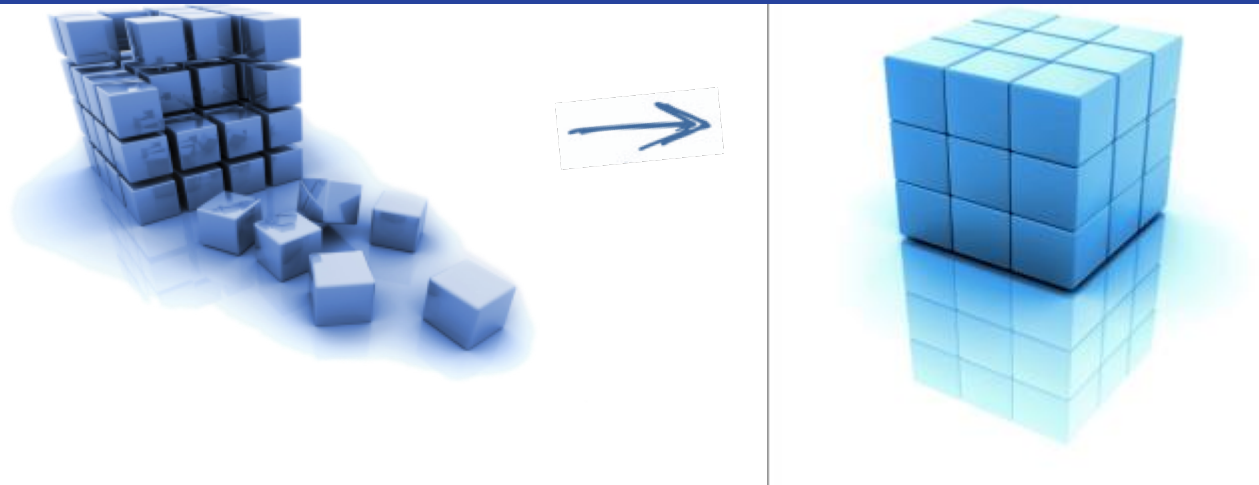


Smart Transformation of Clinical & Nonclinical Data for Insights

Machine Learning and AI augmented transformation of Clinical and Biomarker Data for semantically harmonized access



PhUSE Connect 2019

Smart Transformation

➤ Problem: Disparate data

- ✓ Inconsistent data models and terminologies
- ✓ Unstructured descriptions from studies
- ✓ Challenging to search across studies

➤ Solution: Smart Transformation

- ✓ Artificial intelligence and machine learning augmented transformation of raw as-collected data from data lakes into standardized, usable, accessible, searchable repositories

Smart Transformation

➤ Purpose

- ✓ Curate by looking up Data Lake through Search
- ✓ Transformation Facility to automatically identify re-usable data from raw, legacy or incoming sources.
- ✓ Conversion recommendations of selected data from batches of studies into Structured/Standard format for searchable repository

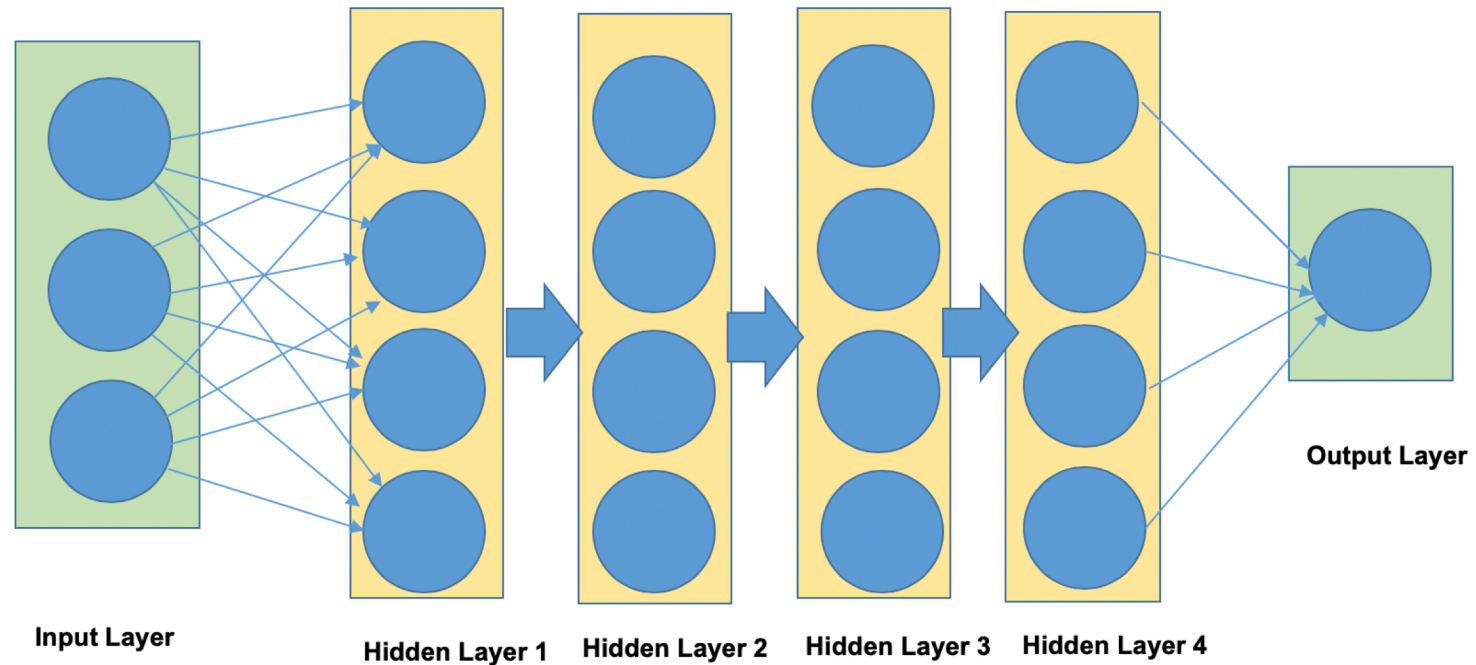
➤ Benefits for Users

- ✓ AI machine learning algorithms use Training sets (Data) & Terminology Governance to recommend mappings
- ✓ “Exception based decisions” for Data Managers to save time
- ✓ Data Validation and QC dashboards to review Data Consistency, Data Formats & Terminology issues

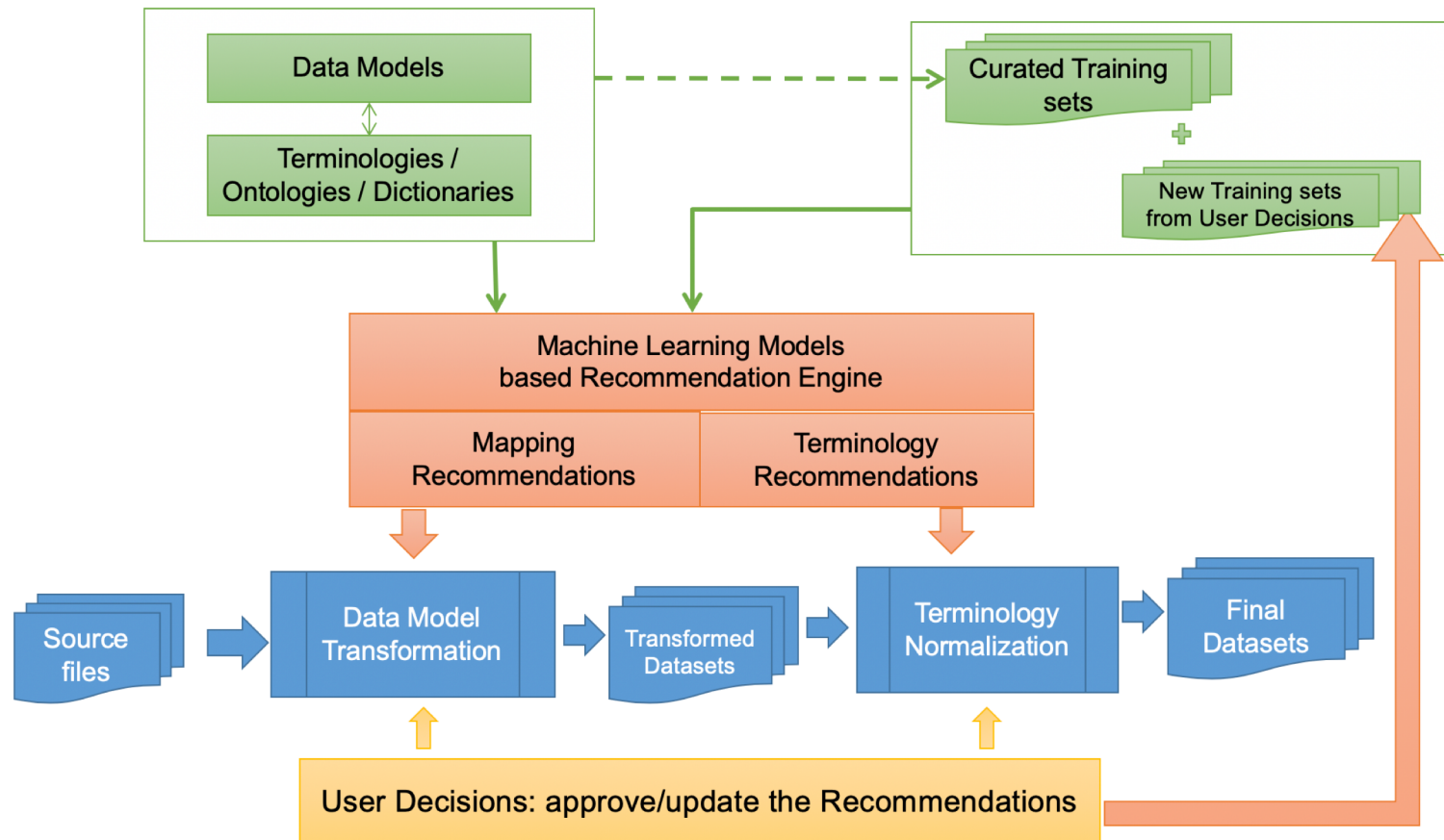
Machine Learning Algorithms

- Artificial Neural Networks based Machine Learning Algorithms
- Pre-trained with Clinical, Non-Clinical and Biomarkers data, and continually learns based on user applied mapping and transformation decisions
- Recommendation algorithms for data mapping, transformation functions and data pivots.
- Recommendation algorithms for Terminology Normalization based on many industry standard databases – CDISC CT, MedDRA, WHODD, GeneInfo, UniProt, miRBase,...

Artificial Neural Networks for Smart Transformation



Smart Transformation Process Flow



Mapping Recommendation (Target Domains & Variables)

Source file metadata

Column Name	Label
PTNO	Patient Number
SEX	Sex
DOB	Date of Birth
INVSITE	Site



Mapping
Recommendations



Recommended mappings

PTNO → DM.SUBJID
SEX → DM.SEX
DOB → DM.BRTHDTC
INVSITE → DM.SITEID

Source data file

PARAM	MATRIX
Basophils	Blood
Eosinophils	Blood
Leukocytes	Blood
Glucose	Urine



Mapping
Recommendations



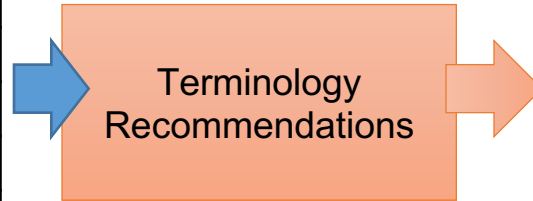
Recommended mappings

PARAM → LB.LBTEST
MATRIX → LB.LBSPEC

Terminology Recommendation (Target Terms)

Transformed dataset

LBSPEC	LBTEST	LBORRESU
Urine	CRE	
Blood	WBC	
	UCREA	
	UGLUCR	
Blood	BASO	G/L
Blood	BASO	%



Recommended Terms

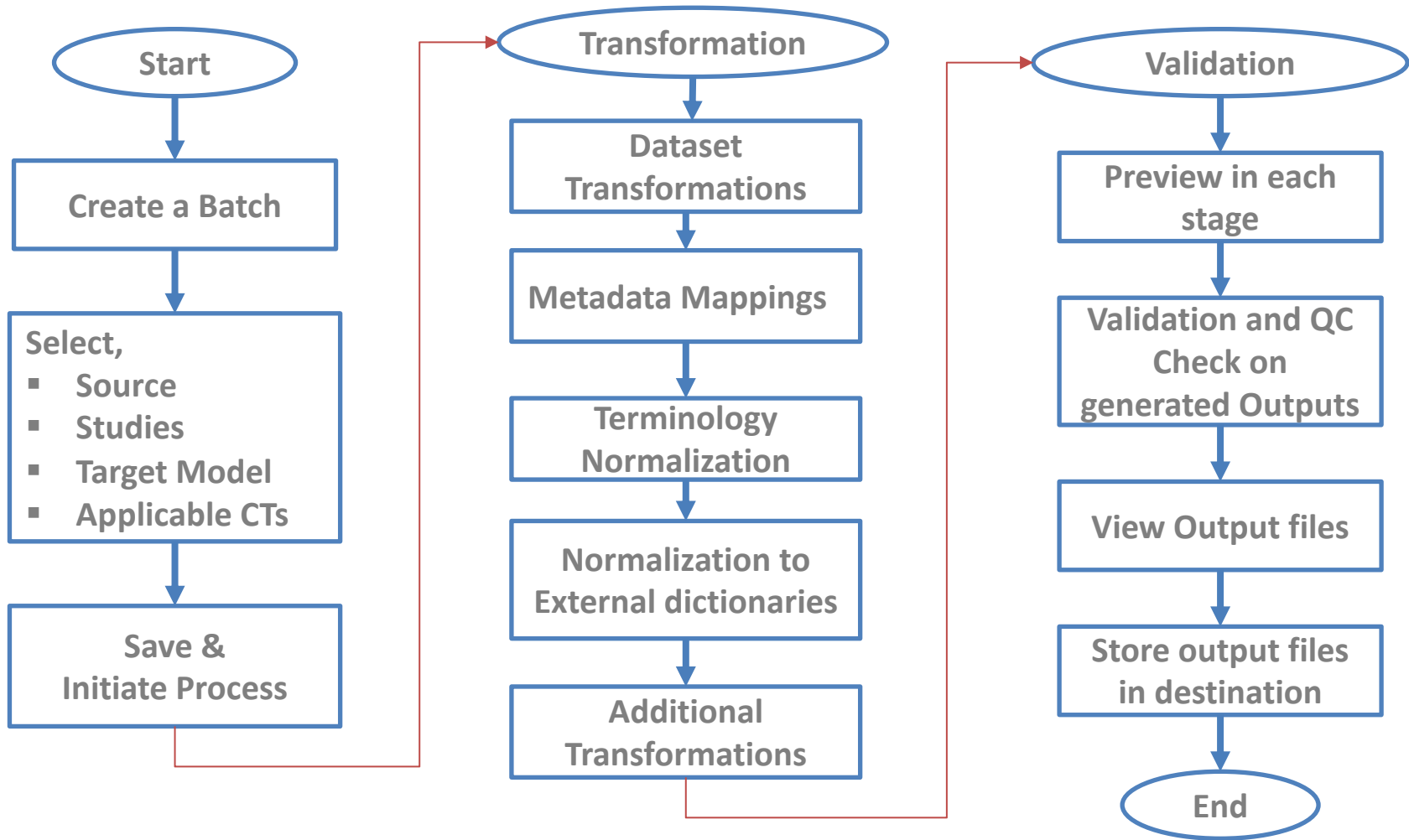
LBSPEC	LBTESTCD	LBTEST
Urine	CREAT	Creatinine
Blood	WBC	Leukocytes
Urine	CREAT	Creatinine
Urine	GLUCCRT	Glucose/Creatinine
Blood	BASO	Basophils
Blood	BASOLE	Basophils/Leukocytes

Smart Transformation

- Recommendation Engine
 - ✓ Created using TensorFlow™ and Keras™
 - ✓ Models trained to classify data domains and variables
 - ✓ Classification results are summarized with a confidence score
 - ✓ User reviews results and can change if necessary
- Continuous learning
 - ✓ User changes tracked
 - ✓ Training datasets updated to capture user input
- Automation to reduce time and effort
 - ✓ Identifying needed data
 - ✓ Semantically mapping and transforming to target format

SUPPLEMENTARY SLIDES

Smart Transformation – High level steps



Mapping Recommendation (Target Domains & Variables)

Metadata Mappings

> **ae** 11 113 1 Target Domain: AE

basecond 9 20 1 Target Domain: MH

#	Source	Mapping	Target	Recommendation
1	ACTEVENT	→	VISITNUM	MH 58% AE 26% more
6	CDMX	→	MHTERM 43% AETERM 38% more	Machine Recommended
9	CPEVENT	→	VISIT	

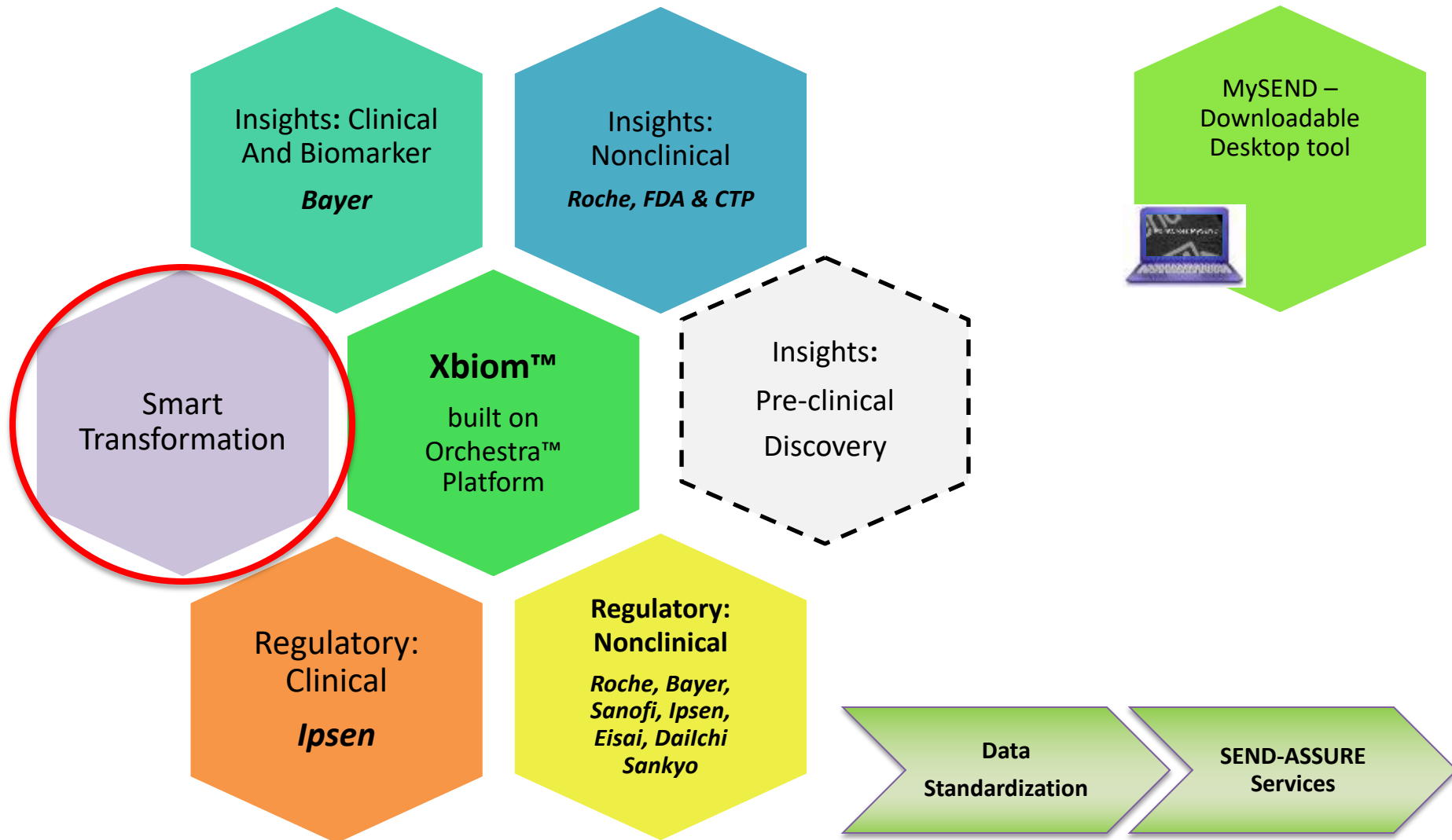
> Preview Choose Study: PC56789

Source file “basecond” data is recommended to “Medical History (MH)” with high confidence.
Source columns CPEVENT & ACTEVENT are recommended to VISIT values.
Based on the data, CDMX column is recommended as MHTERM & AETERM

Terminology Recommendation (Target Terms)

LB				
	Source ↑	Target	Expression	Recommendation
>	Activated Partial Thromb...	Partial Thromboplastin Ti...		
>	Albumin	Albumin Measurement		
>	Alkaline Phosphatase	Alkaline Phosphatase Me...		
>	Bacteria	Bacterial Count		
>	Bilirubin	Total Bilirubin Measurem...		
>	Blood	Blood		
>	Casts	Cast Present Or Absent		
>	Crystals	Crystal Present Or Absent		
>	Epithelial Cells	Epithelial Cell Count		
>				

Xbiom[©] Suite for Research & Regulatory in Clinical & Nonclinical



Xbiom – Technology Stack

Web Server



RDBMS



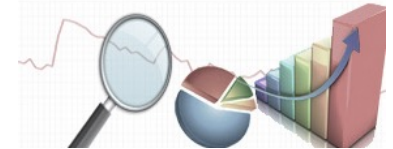
Bigdata Processing



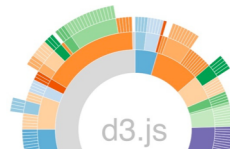
Normalized data
big table



Data Analysis,
Interaction and Search



presto



Orchestra™

Single Sign-on
Web Services
HTTPS

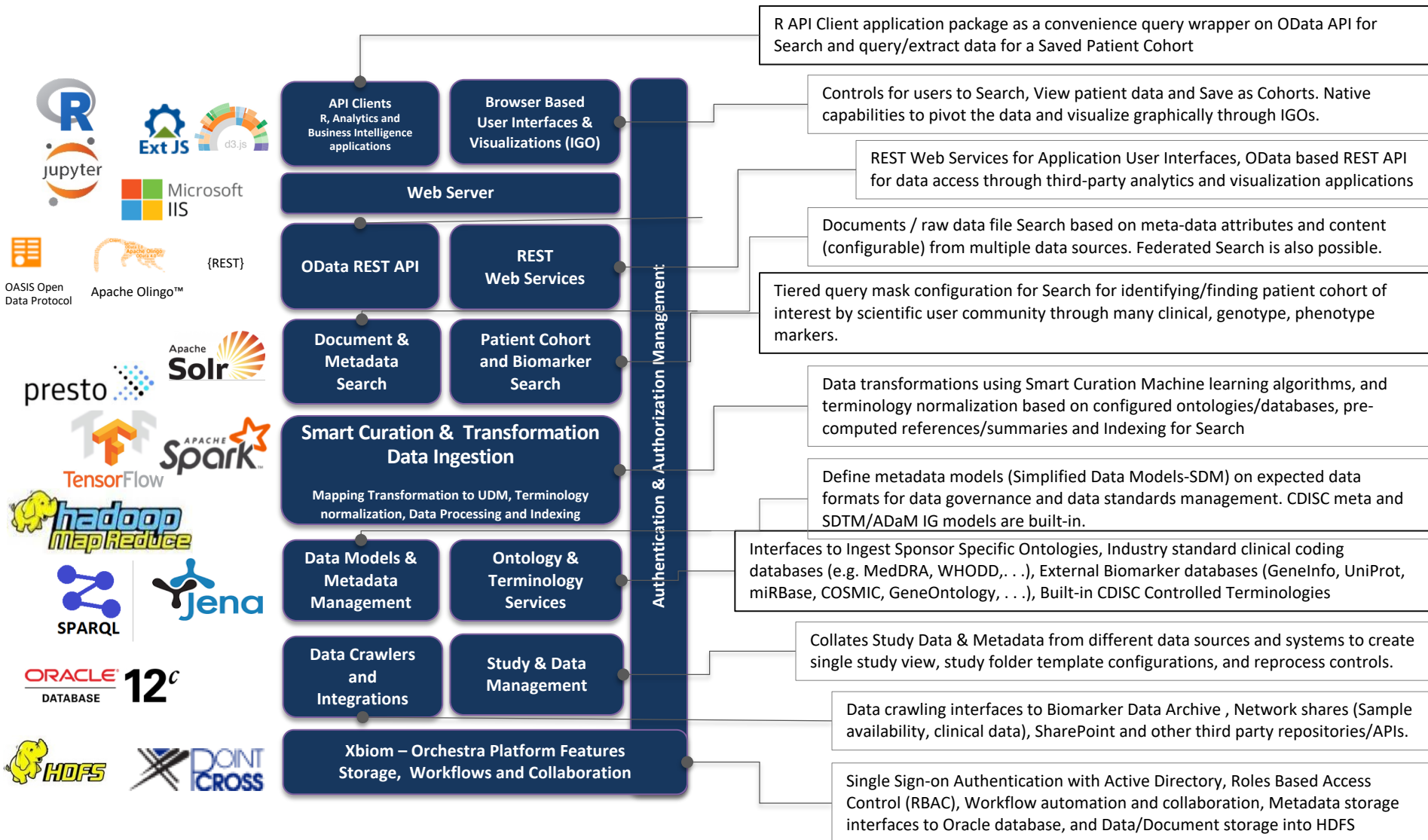
User Accounts & Roles
Object meta-data
Crawler configurations
And Mappers

Smart Curation with Machine
Learning Algorithms for Data
Standardization and
normalization.

Normalized data with
Precompiled Summaries and
statistics

Data visualization, Search and
interactivity

Xbiom - Technical Architecture



*All product names, logos, and brands are property of their respective owners. All company, product and service names used in here are for identification purposes only. Use of these names, logos, and brands does not imply endorsement.

THANK YOU

Isaac Mativo < isaac@pointcross.com >
get_started@pointcross.com

PointCross Life Sciences
www.pointcrosslifesciences.com