Universal Data Model (UDM) for a Streamlined Pathway to SDTM and ADaM Standards

Sunil Gupta, GuptaProgramming@gmail.com
Clinical Data Consultant to PointCross

Rahul Madhavan, Rahul@PointCross.com
VP – Strategic Programs, PointCross

Leverage a Universal Data Model (UDM) to Streamline Pathways to SDTM and ADaM Standards

- Paradigm Shift
  - o Post CDISC Standards and Experience
  - o Non-Linear Accretive Processing
  - o Integrated CDISC Compliance
- Xbiom with Universal Data Model (UDM)
  - o Low-Code User Interface
  - o Six Step Visualization Process
- Universal Data Model (UDM)
  - o Reusability and Repeatability
  - o Direct to Review and Analysis
  - o Monitoring and Accretive Resolution of Data Issues
- SDTM Generation
  - • SDTM IG Specifications, CDISC 360's Mission
  - • Metadata Repository, Auto-Mapping and User Confirmation
  - • Continuous Learning Process

UDM

> **Search and Query Interactive Database**
> **Ingest, Catalog, Index and Link All Data**

Paradigm Shift: Post CDISC Standards and Experience

CDISC Standards / Metadata

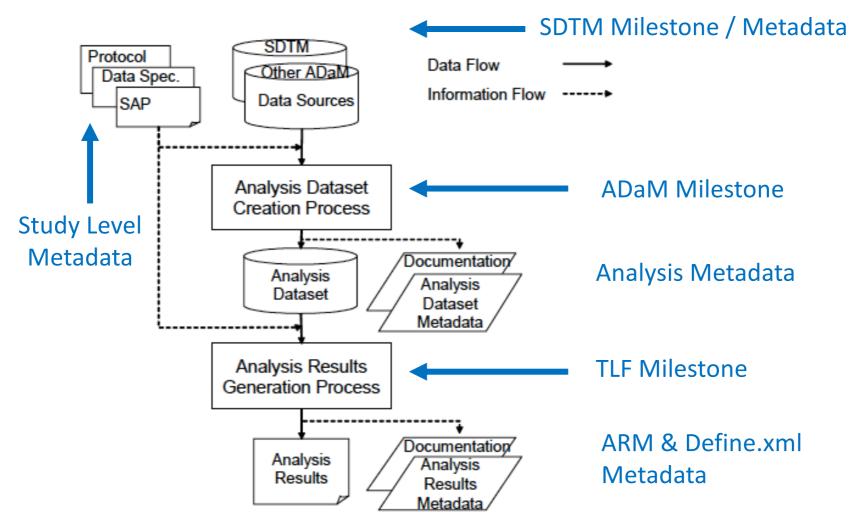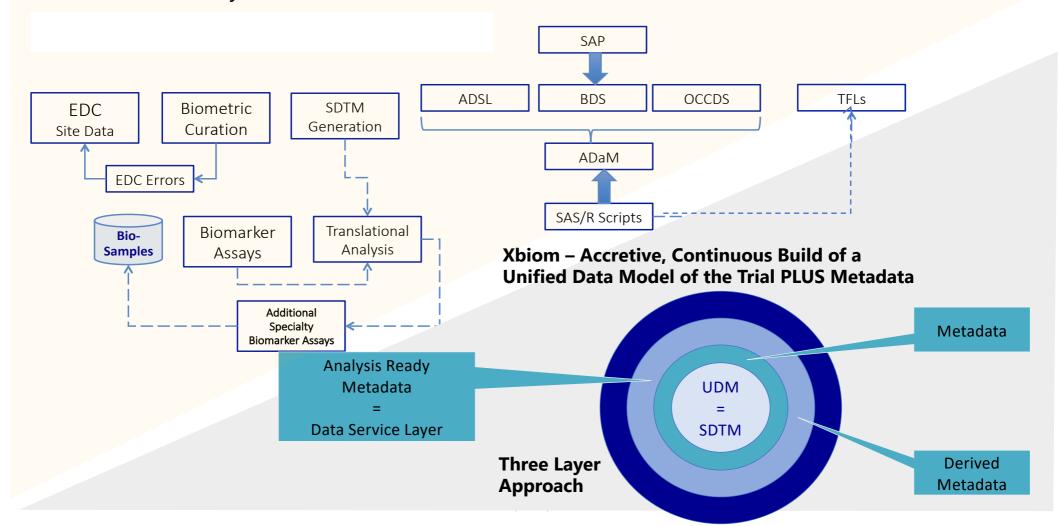Figure 3.2.1: Analysis Data Flow Diagram Showing One Scenario for the Flow of Data and Information

Paradigm Shift: Non-Linear Accretive Processing
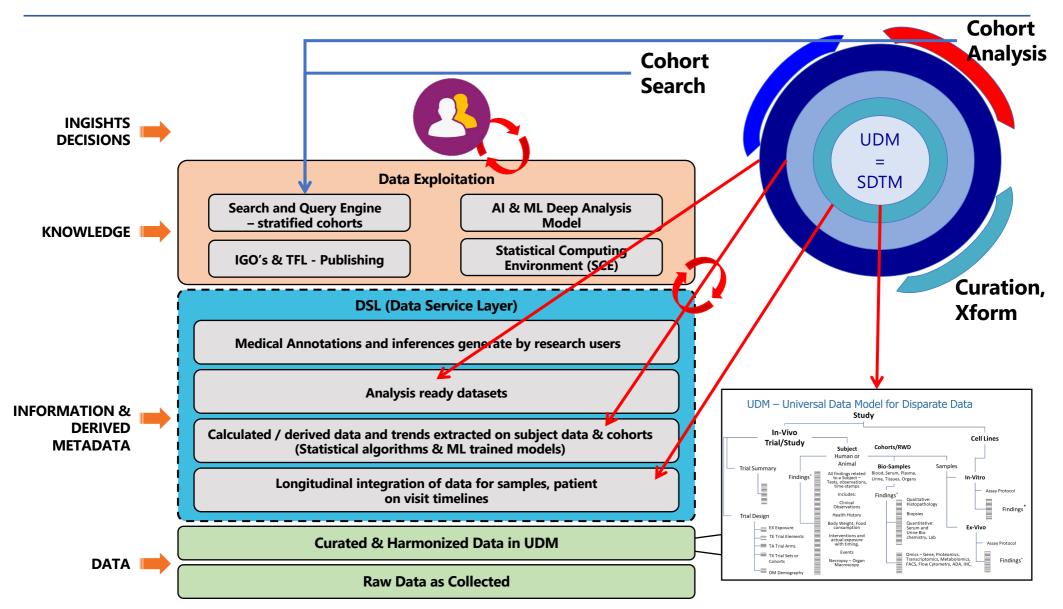
POINT **CROSS** Life Sciences

From Sequential Processing of Files to Create SDTMs & ADaM for TLFs
To A Single Accretivel`y Curated Unified Data Model - Select Cohorts, Analyze Data to SAP,
and Automatically Generate SDTM, ADaM for Submission

SAP

ADSL | BDS | OCCDS | TFLs

EDC Site Data | Biometric Curation | SDTM Generation

EDC Errors

ADaM

Bio-Samples | Biomarker Assays | Translational Analysis

SAS/R Scripts

Xbiom – Accretive, Continuous Build of a Unified Data Model of the Trial PLUS Metadata

Additional Specialty Biomarker Assays

Analysis Ready Metadata = Data Service Layer

UDM = SDTM

Metadata

Derived Metadata

Three Layer Approach

Universal Data Model (UDM) with Data Service Layers

INGISHTS DECISIONS

Cohort Search

Cohort Analysis

KNOWLEDGE

Data Exploitation

- Search and Query Engine – stratified cohorts
- AI & ML Deep Analysis Model
- IGO's & TFL - Publishing
- Statistical Computing Environment (SCE)

UDM = SDTM

Curation, Xform

DSL (Data Service Layer)

- Medical Annotations and inferences generate by research users
- Analysis ready datasets
- Calculated / derived data and trends extracted on subject data & cohorts (Statistical algorithms & ML trained models)
- Longitudinal integration of data for samples, patient on visit timelines

INFORMATION & DERIVED METADATA

DATA

- Curated & Harmonized Data in UDM
- Raw Data as Collected

UDM – Universal Data Model for Disparate Data

Study

In-Vivo Trial/Study

Subject Human or Animal

Cohorts/RWD

Cell Lines

Trial Summary

Findings*

All findings related to a Subject – Tests, observations, time-stamps

Bio-Samples Blood, Serum, Plasma, Urine, Tissues, Organs

Samples

In-Vitro

Includes:
Clinical Observations

Findings*

Qualitative: Histopathology

Assay Protocol

Trial Design

Health History

Biopsies

Findings*

EX Exposure

Body Weight, Food consumption

Quantitative: Serum and Urine Bio-chemistry, Lab

Ex-Vivo

TE Trial Elements

Interventions and actual exposure with timing,

TA Trial Arms

Events

TX Trial Sets or Cohorts

Necropsy – Organ Macroscopy

Omics – Gene, Proteomics, Transcriptomics, Metabolomics, FACS, Flow Cytometry, ADA, IHC,

Assay Protocol

Findings*
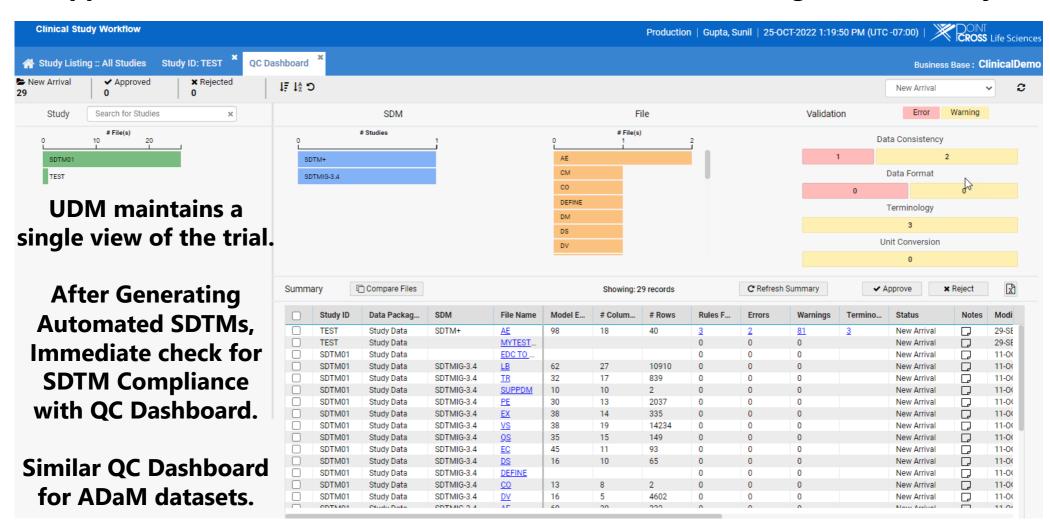
DM Demography

CDISC Compliance: SDTMs and ADaMs

✓ **Traceability**: SDTMs and ADaMs (variables, value level metadata) traced back to SDTM, raw source data, and mapping decisions.

✓ **Reproducibility**: SDTMs and ADaMs may be reproduced using mapping programs. TLF can be reproduced using TLF programs.

✓ **Conformance**: SDTM and ADaM dataset are CDISC conformant for exchangeability i.e. they meet the specific design specifications as well as the general SDTM and ADaM model specifications.

✓ **Completeness**: SDTM are complete, i.e. all raw data items that were designed to be mapped are indeed present in the SDTM datasets – not just EDC data. ADaMs are complete, I.e. all analysis variables are mapped from SDTMs.

✓ **Integrity**: Data integrity is preserved, i.e. data points are not inadvertently affected (e.g. by truncation), no loss of records.

Multiple QC Dashboards Integrated Data & Work Flows

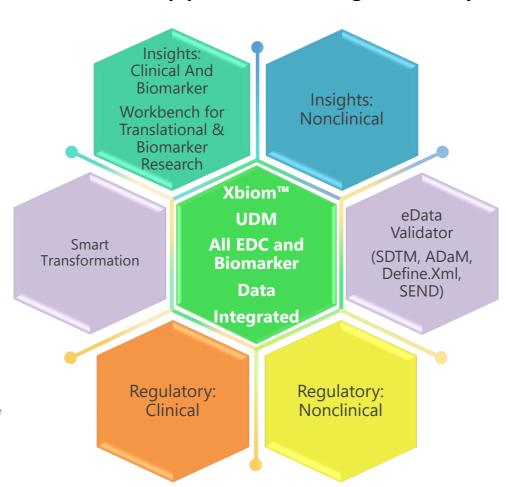## Supports Curation, Transformation, Conformance, Trial Design, Data Quality

UDM maintains a single view of the trial.

After Generating Automated SDTMs, Immediate check for SDTM Compliance with QC Dashboard.

Similar QC Dashboard for ADaM datasets.

Xbiom: Low-Code User Interface and Visualization Tool

PointCROSS Life Sciences

Xbiom is a one-stop platform for integrated study data

Capability

> **Insights for Translational Research in Monitoring Ongoing Studies/Trials**

> **Data Curation, Transformation, Harmonization, Quality, and Ingestion**

> **Regulatory Data Standardization, Submission Planning and Preparation**

Insights: Clinical And Biomarker Workbench for Translational & Biomarker Research

Insights: Nonclinical

Smart Transformation

Xbiom™ UDM All EDC and Biomarker Data Integrated

eData Validator (SDTM, ADaM, Define.Xml, SEND)

Regulatory: Clinical

Regulatory: Nonclinical

R&D Business Value

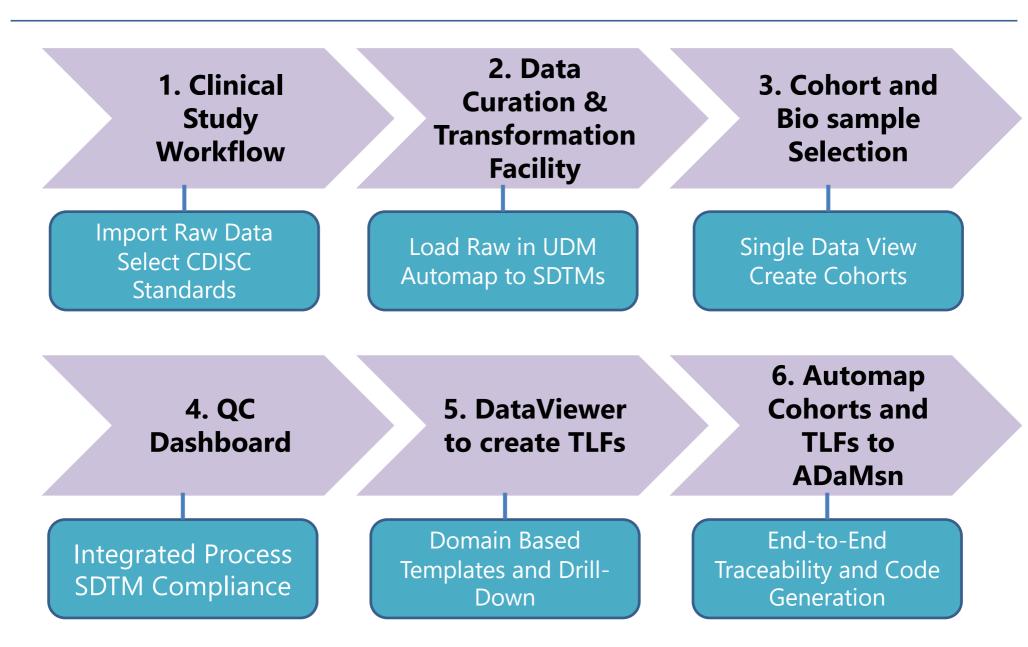Speed to reliable, analysable, and submittable data

Search, define and analyse stratified cohorts

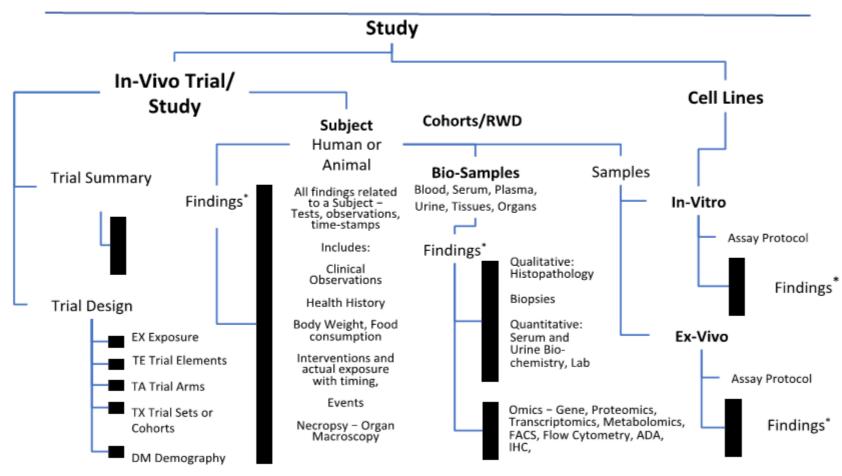Longitudinally integrated subject and patent Data

Rapidly generate TFLs for publication and collaboration

The Xbiom Six Step Visualization Process

1. Clinical Study Workflow

Import Raw Data Select CDISC Standards

2. Data Curation & Transformation Facility

Load Raw in UDM Automap to SDTMs

3. Cohort and Bio sample Selection

Single Data View Create Cohorts

4. QC Dashboard

Integrated Process SDTM Compliance

5. DataViewer to create TLFs

Domain Based Templates and Drill-Down

6. Automap Cohorts and TLFs to ADaMsn

End-to-End Traceability and Code Generation

Unified Data Model: Reusability and Repeatability, Direct to Review & Analysis, Monitor and Accretive Resolution of Data Issues

POINT CROSS Life Sciences

UDM – Universal Data Model for Disparate Data

Study

In-Vivo Trial/Study

Cell Lines

Subject Human or Animal

Cohorts/RWD

Trial Summary

Findings*

Bio-Samples Blood, Serum, Plasma, Urine, Tissues, Organs

Samples

In-Vitro

Trial Design

All findings related to a Subject – Tests, observations, time-stamps

Assay Protocol

Findings*

Includes:

Findings*

Qualitative: Histopathology

Findings*

- EX Exposure
- TE Trial Elements
- TA Trial Arms
- TX Trial Sets or Cohorts
- DM Demography

Clinical Observations

Biopsies

Health History

Quantitative: Serum and Urine Bio-chemistry, Lab

Ex-Vivo

Body Weight, Food consumption

Interventions and actual exposure with timing,

Assay Protocol

Events

Omics – Gene, Proteomics, Transcriptomics, Metabolomics, FACS, Flow Cytometry, ADA, IHC,

Findings*

Necropsy – Organ Macroscopy

* Findings are held in a common storage model with Observation/Testname, date-Time stamp, Units and other attributes

SDTM Automation By Reusability and Repeatability

CDISC 360: Seeks to demonstrate how standards enable metadata-driven end-to-end automation.

Replication and Automation
[Metadata and Data,
Mapping and Transformation, Code Generation]

Metadata-Driven Process and Intelligence
[Global, Study Level, Raw Data (EDC, PK),
Data Meaningful, Machine Readable]

CDISC Metadata and Rules
[Standards, Versions, Data Exchange]

EVOLUTION

Metadata Categories and Examples Mapped to Xbiom

POINT CROSS Life Sciences

SAS and R Programs/Applications: Macro variables, parameters, defaults, programs and libraries, functions, conditional macro calls and assignments (If/Then/Else), operators, templates, styles, applications, data-driven macros, lookup tables, design/logic, code generator, user message, libnames, config file

Context: Meaning, Purpose, mind map

CDISC Model and Compliance: Rules, Events/Findings, structure (BDS/Wide), Order, codelist, define.xml (hierarchy), standard/original values, map/unmap variables
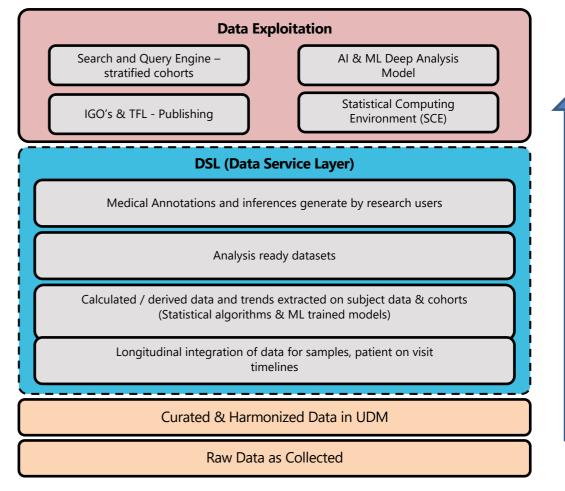
Dataset and Variable Relationships: 1:M, Parent: Child, top 10, links

Folders, Files, Dataset, Variables Attributes: length, type, format, color, size, weight, access, location, units

Folders, File, Dataset and Variable Names: lists, file name, extension type
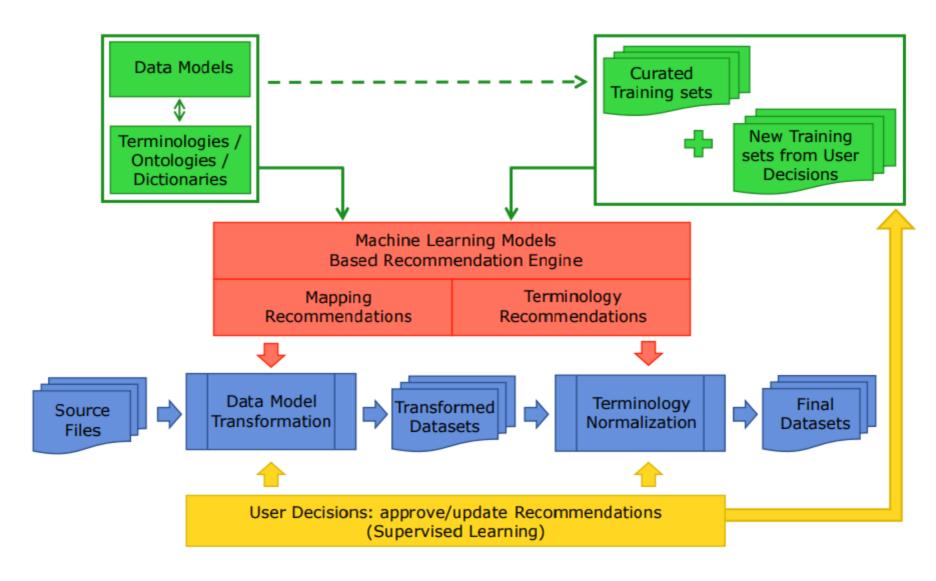
Variable Values: valid values, code list, min, max, continuous, categorial, format catalog, statistical analysis, cutoff/flags, new/change, date, times, patterns

Data Exploitation

- Search and Query Engine – stratified cohorts
- AI & ML Deep Analysis Model
- IGO's & TFL - Publishing
- Statistical Computing Environment (SCE)

DSL (Data Service Layer)

- Medical Annotations and inferences generate by research users
- Analysis ready datasets
- Calculated / derived data and trends extracted on subject data & cohorts (Statistical algorithms & ML trained models)
- Longitudinal integration of data for samples, patient on visit timelines

Curated & Harmonized Data in UDM

Raw Data as Collected

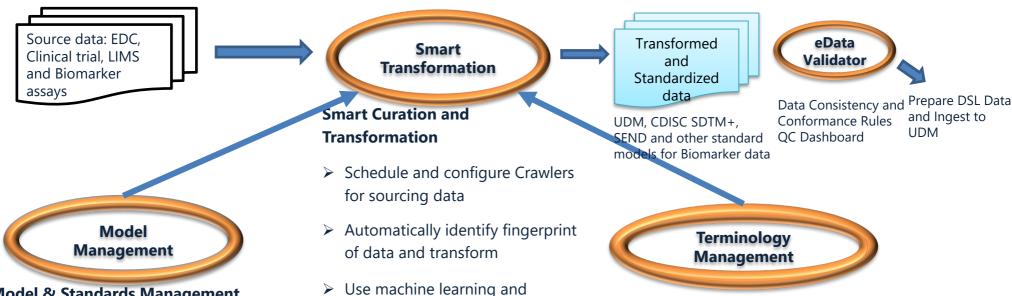EVOLUTION

Machine Learning and Data Transformation Work Flow

Data Curation is the Repetitive Process to Optimize Data and Metadata to ensure Valuable use of Data.

Smart Curation and Transformation

POINT **CROSS** Life Sciences

Source data: EDC, Clinical trial, LIMS and Biomarker assays

Smart Transformation

Transformed and Standardized data

UDM, CDISC SDTM+, SEND and other standard models for Biomarker data

eData Validator

Data Consistency and Conformance Rules QC Dashboard

Prepare DSL Data and Ingest to UDM

Smart Curation and Transformation

➤ Schedule and configure Crawlers for sourcing data

➤ Automatically identify fingerprint of data and transform

➤ Use machine learning and supervised teaching to continuously improve automation and quality of curation and transformation

➤ Re-use and specify Transformation target Data model and Terminologies

➤ transform Source data to target models for ingestion

Model Management

Model & Standards Management

➤ Generate specifications for data model and validation based on DTS/DTA

➤ Build Enterprise's proprietary models

➤ Maintain model registry for wide re-use of corporate supported standards

➤ Import standards and latest versions (e.g. CDISC SHARE) in source, XML, or RDF

Terminology Management

Ontology ./ Terminology Management

➤ Maintain ontology of all global and local terminologies for wide re-use

➤ Build Enterprise's internal or local terminology lists (for "preferred use")

➤ Import Controlled Terminology and code-lists from multiple global public or subscription registries (MedDRA, SNOMED, CDISC, WHODD, genomic registries, NCI, INHAND, …)

Metadata Repository: Conceptual Data Model (UDM)
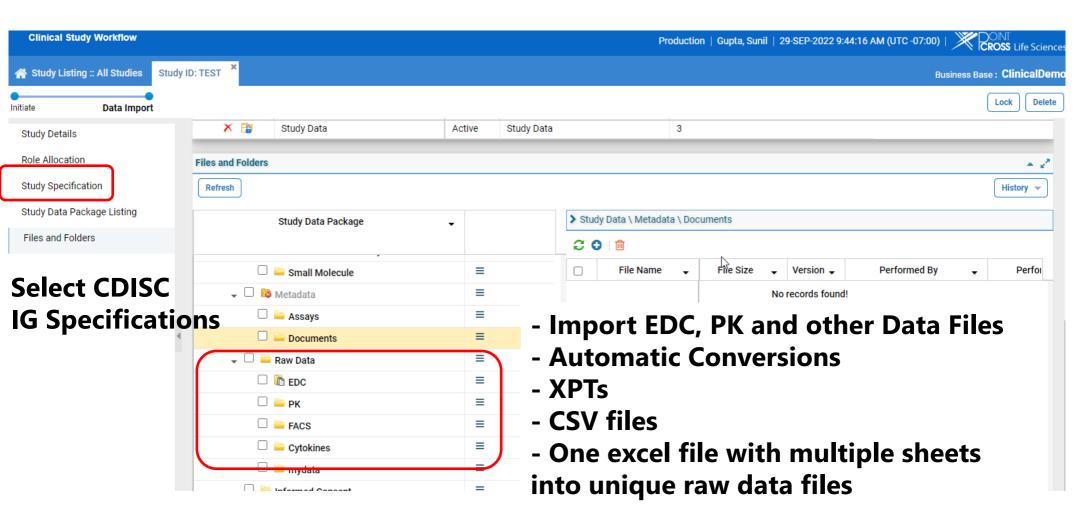
Simple, Extensible and Ready for analytics

Compound / Study metadata

Study

Compound Information

Trial design data
Interventions data
Documents

Statistical Analysis Plan / Reports / Documents

Trail Design Data (Arms/Elements)

Interventions Data

Subjects metadata

Subjects

Findings data

BioSample Availability

Informed Consent

Imaging, High-dimensional Genomic raw datasets

Molecular biomarkers (Omics & External Lab)

Lab Findings /Events Data

Endpoints / Analysis data

Pharmaco Vigilance

EMR / EHR

Taxonomy Lists & Biomarker DBs
(Local: Master data, Standard: CDISC SDTM, SEND, ADaM, MedDRA, WHO Drug, Gene Info, UniProt...)

Ontologies, Dictionaries
Master data / External Databases

Metadata Example: Analysis Results

Metadata Field	Metadata
DISPLAY IDENTIFIER	Table 12.3.1.1
DISPLAY NAME	Mean NRS Pain Score Over the Last 5 Days for Overall Pain. Full Analys Set
RESULT IDENTIFIER	Treatment difference results (Mean, confidence interval, p-value)
PARAM	Overall Pain Score during the 5-day Period
PARAMCD	PLPNOV
ANALYSIS VARIABLE	CHG, BASE, TRT02AN, GEOREGN
REASON	Primary efficacy analysis as pre-specified in protocol
DATASET	ADQS
SELECTION CRITERIA	fas1fl='Y', paramcd='PLPNOV', trt01pn~=., avisit='EoT'
DOCUMENTATION	See Protocol Section XX for details. Program: program_ex1.sas. NRS sco were analysed using an ANCOVA model which included dose group and region (REG1 and REG2) as fixed factors and baseline NRS pain score of overall pain a covariate.
| PROGRAMMING STATEMENTS | ```
data pain;
    set adam.adqs;
    where fas1fl='Y' and paramcd="PLPNOV" and
avisit="EoT";
run;
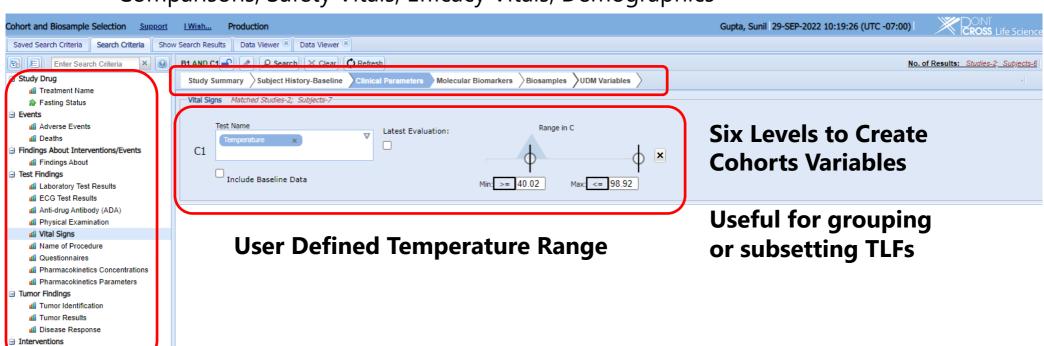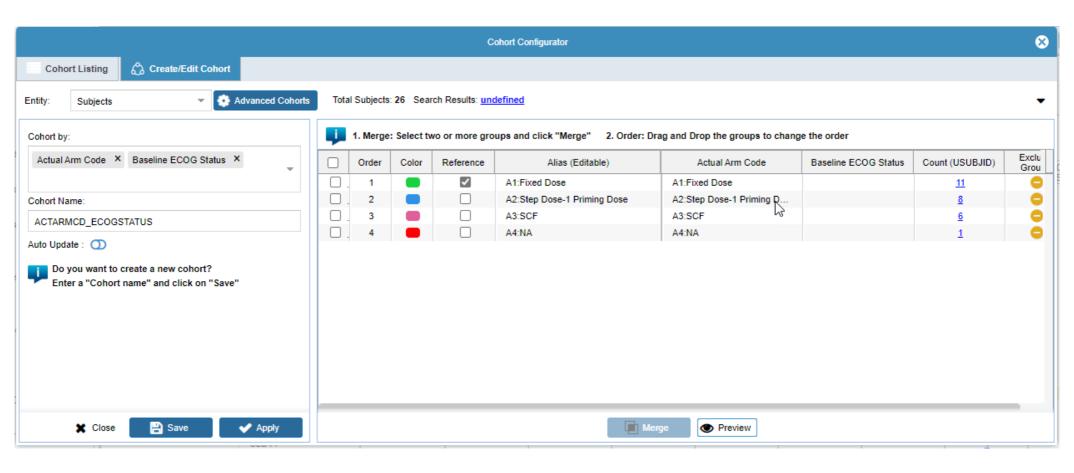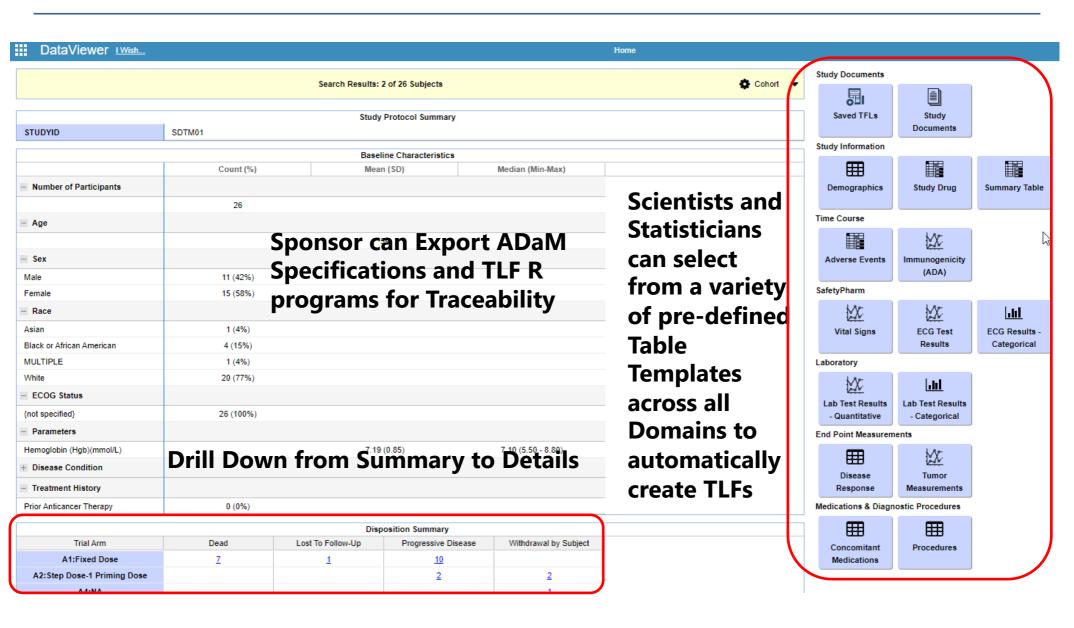
proc mixed data=pain;
    class &trt georegn;
    model chg=base &trt georegn;
    lsmeans &trt/cl adjust=dunnett;
    estimate 'Linear trend' &trt -2 -1 0 1 2;
    ods output type3=pvalue;
    ods output lsmeans=lsm;
    ods output diffs=dif;
    ods output estimates=trend;
run;
``` |

TFL Metadata

TFL Specifications

ADaM Metadata

Protocol /SAP

SAS Code

Clinical Study Workflow: CDISC IG Specs, Raw Data

Select CDISC IG Specifications

- **Import EDC, PK and other Data Files**
- **Automatic Conversions**
- **XPTs**
- **CSV files**
- **One excel file with multiple sheets into unique raw data files**

Direct to Review Cohorts and Biosamples

Create Patient Cohort Query Across Any Domain Variables, Ex. CHOLFLG
- Subject Population: TRT01A='Drug A', SAFFL='Y'
- Safety Population: AEREL='Y', AESER='Y'
- Efficacy Population: PARAMCD='CHOL', VISIT='12 WEEK', PCHG=-30

Data and Visual Cohort Review and TLFs
- Cohort, Subject, Multiple-Endpoints, Summary, Statistics, Cross-Reference, Comparisons, Safety Vitals, Efficacy Vitals, Demographics

Six Levels to Create Cohorts Variables

Useful for grouping or subsetting TLFs

User Defined Temperature Range

Search, Find, Select or Create Cohorts based on SAP or Needs of Translational Research

PointCROSS Life Sciences

Powerful and Flexible Search Conditions from Any Domain

Create Simple or Stratified Cohorts

Used to create ADaM variables

Cohort Configurator

| | Cohort Listing | Create/Edit Cohort |

Entity: Subjects ▼ | ⚙ Advanced Cohorts | Total Subjects: 26 Search Results: underlined

Cohort by:
Actual Arm Code ✕ | Baseline ECOG Status ✕

Cohort Name:
ACTARMCD_ECOGSTATUS

Auto Update : ⬤

ℹ Do you want to create a new cohort?
Enter a "Cohort name" and click on "Save"

1. Merge: Select two or more groups and click "Merge"   2. Order: Drag and Drop the groups to change the order

☐	Order	Color	Reference	Alias (Editable)	Actual Arm Code	Baseline ECOG Status	Count (USUBJID)	Exclu Grou
☐	1	🟢	☑	A1:Fixed Dose	A1:Fixed Dose		11	⊖
☐	2	🔵	☐	A2:Step Dose-1 Priming Dose	A2:Step Dose-1 Priming D...		8	⊖
☐	3	🔴	☐	A3:SCF	A3:SCF		6	⊖
☐	4	🔴	☐	A4:NA	A4:NA		1	⊖

✕ Close | 💾 Save | ✔ Apply

🖥 Merge | 👁 Preview

DataViewer Panel to create Tables for 'End-in-Mind'

POINT CROSS Life Sciences

Sponsor can Export ADaM Specifications and TLF R programs for Traceability

Drill Down from Summary to Details

Scientists and Statisticians can select from a variety of pre-defined Table Templates across all Domains to automatically create TLFs

Clinical Data Issues

- Missing Data Values
- Invalid Dates and Data Values
- Character / Numeric Variable Type Conversion
- Zero Records

Study Protocol Data Issues

- Standardized Terms – no loss of data or context
- Lab Data – identify duplicate records, missing values, invalid units, etc.
- Primary Endpoints – correctly derived
- Survival Analysis – subgroup analysis
- Safety – maximum patients and events
- Deaths – maximum patients
- Related Adverse Events – minimum patients and events
- Protocol Compliance – visit in visit window range

SDTM: Automapping and User Confirmation, Metadata Repository and Continuous Learning Process

POINT CROSS Life Sciences

Biomarker Data
(Specialty Labs)

Periodic Scrape → **1 Day** → Longitudinally Integrated Data Available for Decisions

For SDTM and ADaM Generation

EDC Site Data

Asynch Periodic Scrape (Translational)

Planned Data Lockout

SAP Statistical Analysis Plan

Xbiom™

Automated Curation

UDM Unified Representation of ALL Patient Data

UDM->SDTM Generator (with Define.XML, Trial Design Domains)

Custom ADaM (ADSL, BDS, OCCDS) Generator for SAP Analysis)

Biometrics & Bio-Statistics Analysis (SCE)

TFLs

SDTMs/ADaMs/TLFs
Integrated Metadata Mapping and Repository

POINT **CROSS** Life Sciences

Input

Xbiom

Outputs

EDC Raw

➡

(UDM)
User Interface to Matched
and Un-Matched
Metadata
- Outputs 'At-Will'
- Save Queries
- Datasets, Variables
  , Control Terms
- SDTM/ADaM/TLF
  Templates /
  Compliance Rules
- Dataset Joins
- User Batch Process for
  Data Issues / Data
  Updates
- Flag Data Changes /
  Issues

➡

Standard SDTMs,
ADaMs, TLFs

Biomarker Data

➡

**Based on
SDTM/ADaM IG
Rules**

Other Data,
ex. PK

➡

➡

Custom SDTMs:
SUPPXX, RELREC, FA

SDTM Auto-Mapping and User Confirmation

- Load Raw Data
- Automap Raw Data to SDTMs (UDM)
- User Interface to Confirm Automapping Structures (Variables, Units, Values)
- User to define non-auto mapped Raw Data

SDTM Auto-Mapping and User Confirmation

Machine Process Remembers Auto-Mapping and User Selections as Lookup Tables

Auto-Mapping and Continuous Learning Process

CDISC 360: Apply the 80/20 rule to ensure the Project automates 80% of the end-to-end metadata and data processing needed to generate study artifacts suitable for a regulatory submission.

Overall Process
- Pre-processing Batch
  - Three Variable Mapping Methods
  - Two User Approval Methods
  - Two Control Terms Mapping Methods
  - Variable Derivations
- Data Update Batch

Variable Mapping Methods
1. Direct
2. Transformation, SQL, ex. trim, concatenating
3. Transpose
4. One Raw Data to Multiple SDTMs
5. Multiple Raw Data to One SDTM

User Approval Methods
1. Machine Recommended
2. Previous Decision
3. Preview Raw and SDTM
4. RELREC, SUPPXX, FA

Control Terms Mapping Methods
1. Exact Value Match
2. Approximate Value Match

Variable Derivations
DY, STDY, ENDY, DTC, SDTC, ENDTC, BLFL, VISIT

SDTM Mapping Automation Process Examples

POINT CROSS Life Sciences

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

Transformed dataset

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

→ Terminology Recommendations →

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

POINT CROSS Life Sciences

Automation and User Interface Tool

- **Weeks to write SAS Macros and Programs**
- **Non-unified SDTMs**
- **Manually Cross-Reference Biomarker Data**

Traditional SAS Programming

- **Days with Off-the-Shelf Metadata Repository Solution**
- **User Interface Mapping, Cohorts and TLFs**
- **Output Mapping Specifications**
- **Central Data Curation**
- **Drill Down TLFs**

Universal Data Model (UDM) for a Streamlined Pathway to SDTM and ADaM Standards

Sunil Gupta, GuptaProgramming@gmail.com
Clinical Data Consultant to PointCross

Rahul Madhavan, Rahul@PointCross.com
VP – Strategic Programs, PointCross