February 2023

Automating Drug Discovery Informatics

Talking Points on BioChemUDM, Software Selection/Implementation, Data Lake Tahoe, and Harmony-ML™

Automating Drug Discovery Informatics

BioChemUDM, Software Selection/Implementation, Data Lake Tahoe, and Harmony-ML™

- Background and History
- Implementation: Strengths and Limitations
- Machine Learning … and more!

Synthetic Lethality

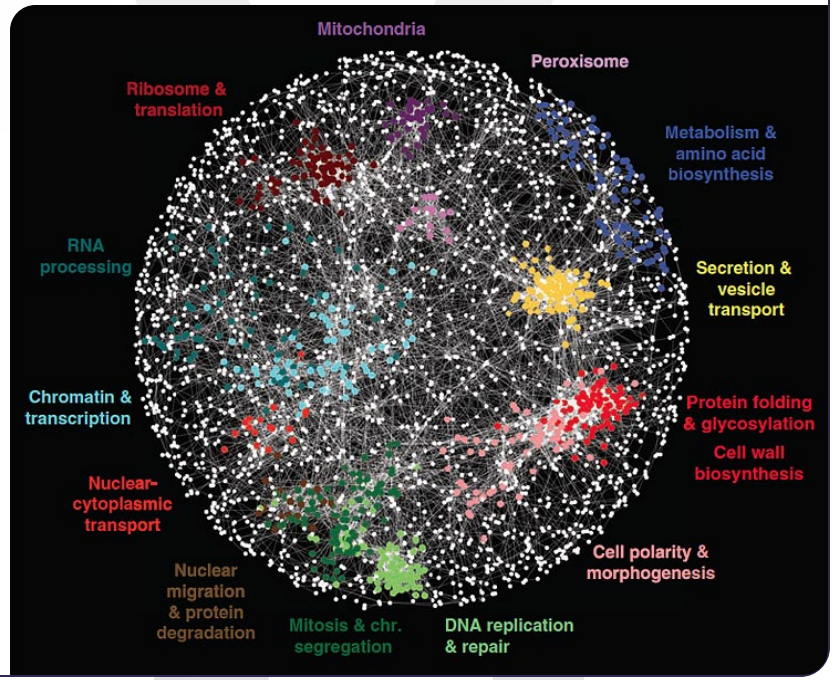
The Next Frontier in Precision Medicine Oncology

Synthetic Lethality provides a powerful approach to discover novel precision medicine therapies with patient biomarkers, including MTAP-deletion (~15% of solid tumors), BRCA/HRD (Breast, Prostate, Ovarian), and high-MSI (15% GI Cancers)

nature REVIEWS GENETICS

- **Synthetic lethality** occurs when the simultaneous perturbation of two genes results in cell death
- Synthetic lethality provides a novel approach to target several historically undruggable loss of function mutations
- Large-scale screening for synthetic lethal targets has progressed through advances in molecular biology (e.g., RNA interference, CRISPR-Cas9 editing) and bioinformatics

Nature Reviews Genetics, Vol. 18, 2017, Hieter, et al., as edited by IDEAYA



Reference: Charles Boone

IDEAYA BIOSCIENCES

IDEAYA's Precision Medicine Oncology Pipeline

Building the Industry Leading Synthetic Lethality Focused Biotechnology Company

Precision Medicine Pipeline

	Modality/Indication	Biomarker	Preclinical	IND Enabling	Phase 1	Phase 2	Program Goals	Collaborations	Commercial (IDEAYA)
Darovasertib *PKC*	+cMET[1] Combination MUM, Basket	GNAQ/11					Daro + Crizo Clincal Update in MUM √ Daro + Crizo Reg Trial in MUM Q1 2023	Pfizer (1)	WW Commercial Rights
	Adjuvant UM	GNAQ/11					(Neo)Adjuvant UM – Phase 1 IST √ Neoadjuvant UM – IDEAYA Phase 1 Q4 2022		
	+cMET[1],+KRAS Combo NSCLC, HCC	MET KRAS					+cMET[3] – Preclinical Evaluation +KRAS – Preclinical Evaluation	Pfizer (1)	
IDE397 *MAT2A*	Monotherapy NSCLC, Esophagogastric	MTAP					Mono Expansion Phase 2 Initiation √		WW Commercial Rights
	Combinations Solid Tumors	MTAP					Combination Cohorts Ph1 Initiation +Pemetrexed√ +Taxanes√ +AMG 193	AMGEN (2)	
IDE161 *PARG*	Ovarian, Gastric, Breast Cancers	HRD					IND Q4 2022	CANCER RESEARCH UK (3)	WW Commercial Rights
Pol Theta	Small Molecule Helicase Inhibitor	HRD					Development Candidate Q2 2022 √ First-in-Human Studies H1 2023	GSK (4)	Global Royalties
WRN	GI Cancers	High-MSI					Development Candidate 2023	GSK (4)	US 50/50 Profit Share Ex-US Royalties
MTAP-SL	Solid Tumors	MTAP					Lead Series		WW Commercial Rights
SL Platform	Solid Tumors	Defined Biomarker					Lead Series New Target / Biomarker Validation		WW Commercial Rights

(1) Pursuant to Pfizer Clinical Trial Collaboration and Supply Agreements for Darovasertib/Crizotinib Combination in MUM and in cMET-driven Tumors; IDEAYA retains all Darovasertib Commercial Rights
(2) Pursuant to Amgen Clinical Trial Collaboration and Supply Agreement for IDE397 + AMG 193, an investigational MTA-cooperative PRMT5 inhibitor; Amgen will sponsor the study and the parties will jointly share external costs of the study
(3) Pursuant to CRUK Evaluation, Option and License Agreement; IDEAYA controls all PARG Commercial Rights
(4) Pursuant to GSK Collaboration, Option and License Agreement: Polθ: Global Royalties; WRN: 50/50 US Profits + ex-US Royalties

= Target Program Milestones

MAT2A=methionine adenosyltransferase 2a, MTAP=methylthioadenosine phosphorylase, MTA=methylthioadenosine, PRMT5=protein arginine methyltransferase 5 (PRMT5), PARG= poly (ADP-ribose) glycohydrolase, DDT = DNA Damage Target, WRN = Werner Helicase, Polθ = DNA Polymerase Theta, HRD = homologous recombination deficiency, MSI = microsatellite instability, PKC = protein kinase C, MUM = metastatic uveal melanoma, cMET = tyrosine kinase protein MET , Crizo = crizotinib, NSCLC = non-small cell lung cancer, HCC= hepatocellular carcinoma WW = worldwide

IDEAYA Synthetic Lethality Drug Discovery Platform

Structure-Based Drug Design & Proprietary Chemical Library Enable "Hard to Drug" Targets

Structural Biology & Structure Based Drug Design

Full suite of capabilities in structural biology, biophysics, & computational chemistry

Ligand bound co-crystal structures resolved to enable Structure Based Drug Design for programs

Multiple potential "first-in-world" co-crystal structures resolved, including for PARG, Pol Theta Helicase and Werner Helicase

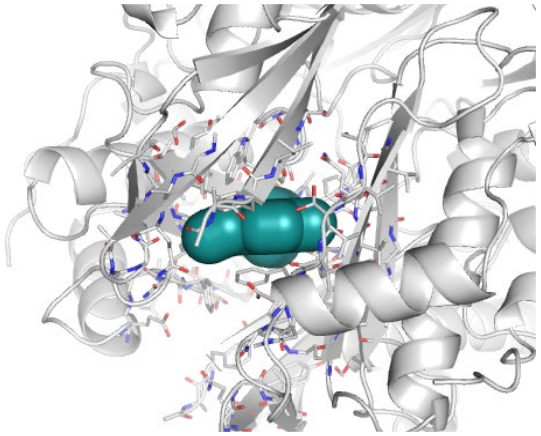
Harmony-ML™ Proprietary Machine-Learning

Our internal ML engine empowers our discovery platform through effective prioritization leading to efficient cycle times

INQUIRE™ Proprietary Chemical Library

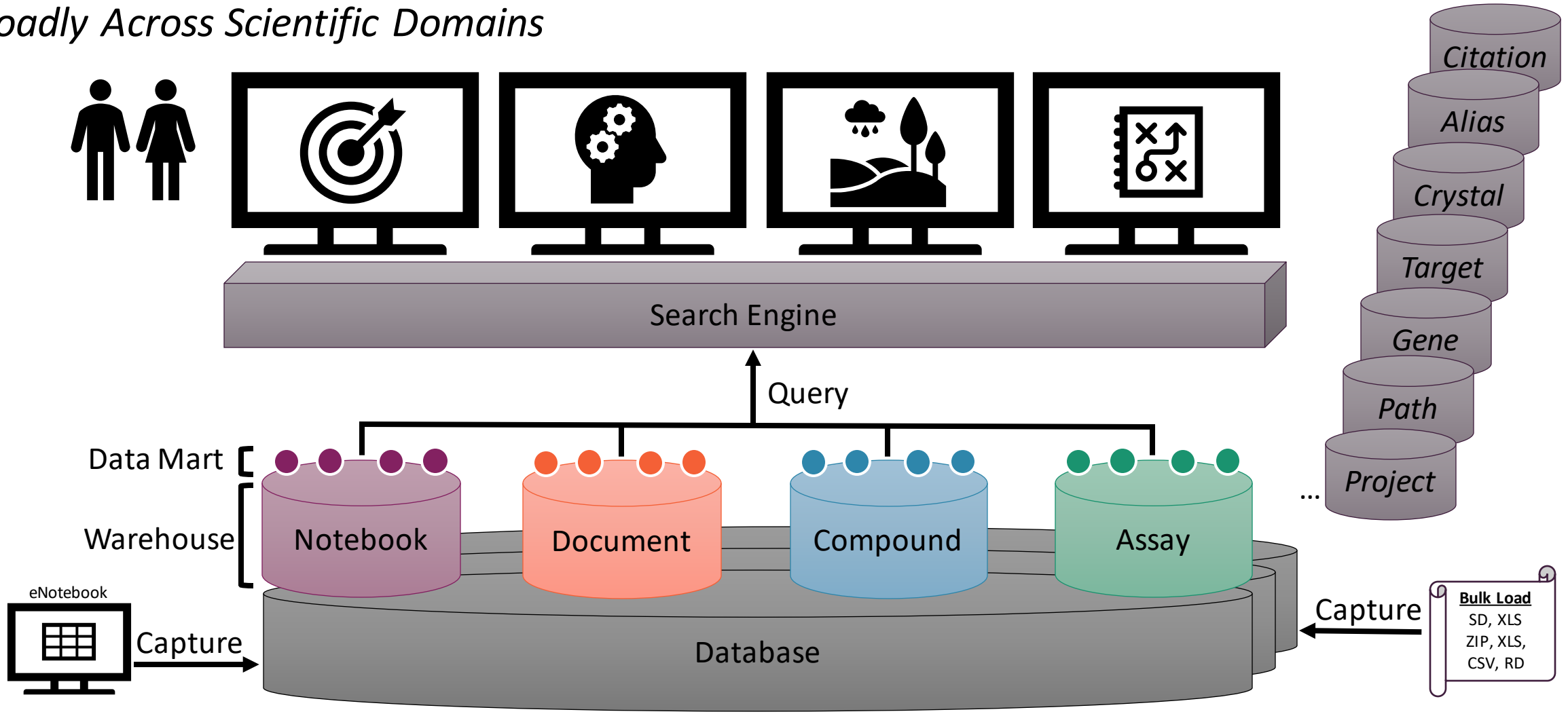
Expert-curated HTS library to enhance hit discovery capabilities against novel SL targets classes, such as helicases and endonucleases

Enhances IDEAYA's SL Drug Discovery Platform and competitive differentiation



The Vision: A Modern Architecture to Leverage Information

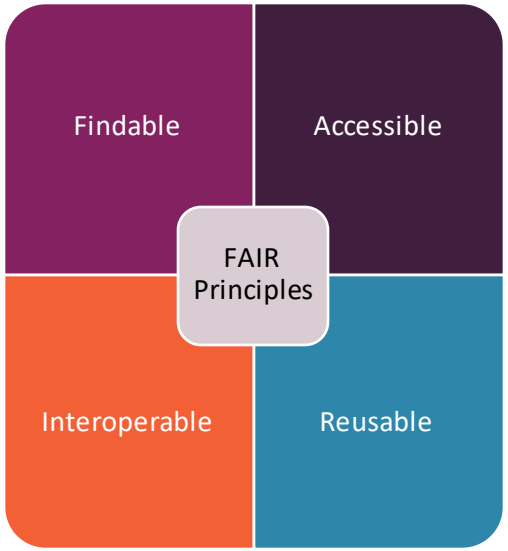
Broadly Across Scientific Domains



Do we need an “Informatics Starter Kit” for drug discovery groups?!?

A Pragmatic Approach to Software Selection/Implementation

- We repeatedly find ourselves registering compounds, defining protocols, capturing assays, visualizing data, and struggling to integrate data from collaborators; a super, simple data model was born, *BioChemUDM*
- By co-adopting the BioChemUDM, we have successfully enabled same-day exchange and utilization of chemical and biological information with various stakeholders
- Implementing FAIR principles with the BioChemUDM, a modern architecture to leverage information broadly across scientific domains can happen quickly without heavy or differentiated lifting



- We *listened* to your needs, *explored* infrastructure, and *requested* proposals for evaluation.
- We have *organized* and *curated* **notebooks, documents, compounds, assays and inventory**.
- We have selected a portion of our data for migration into platforms & visualization tools. (in progress)
- **We need your help!** Our collective evaluation is key to success.
  1. Make a scorecard to evaluate tools that everyone to use. (August)
  2. Evaluate software using the scorecard. (September)
     - Record an experiment in the e-notebook.
     - Registry a compound from an experiment.
     - Capture an assay from an experiment.
     - Search inventory for location and amount.
  3. Select the best to begin the modern architecture. (October)
- Complete curation and migration of all data. (Q4)

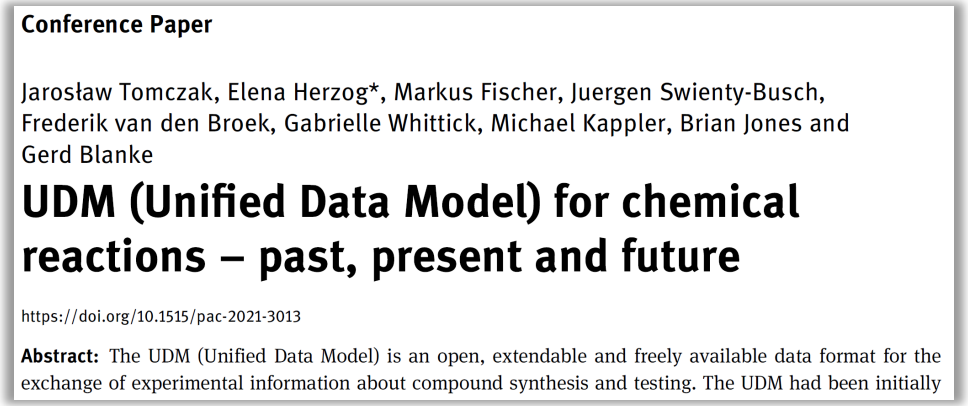
Evaluation Scorecard

	A	B	C	D	E	F
1	CBIS+DW	CDD+SD	DS+SP	SCI+D360	Points	Requirement
2		0		0	100	**Overall**
3			0			
4		0		0	25	**eNotebook**
5			0		3	Use ChemDraw as a sketcher
6				0	2	Attach a document to an experiment
7					2	Update experiment with undesired product
8					1	Register a batch from an experiment
9					1	Capture assay data for a compound sample
10					1	Capture a file with assay data
11					5	Curve fit based on four-parameter logistic model
12					3	Automatic initialization of parameters
13					3	Automatic suggestion of outliers
14					3	Automatic determination of inflection point
15					1	Automatic calculation of standard error
16						Automatic filling of stoichiometry table
17						Text formatting
18						Reagent lookup by name/structure search
19						Automatic

Unified Data Models in Drug Discovery

A Brief History

- *RxnUDM – integrated and intuitive browsing of molecules, reactions, and citations*
  - In 2012, four private electronic notebooks, multiple public reaction databases, two massive compound registries, and references supporting everything merged into Elsevier Reaxys™
  - In 2014, Roche releases RxnUDM to publisher
  - In 2016, Elsevier promotes the UDM into industry
  - In 2018, Pistoia Alliance produces a UDM for public use
  - In 2020, authors collaborate on implementation and adoption



Conference Paper

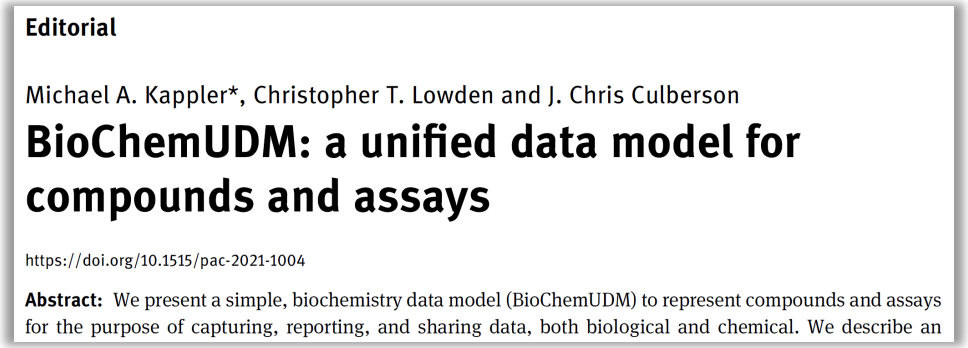
Jarosław Tomczak, Elena Herzog*, Markus Fischer, Juergen Swienty-Busch, Frederik van den Broek, Gabrielle Whittick, Michael Kappler, Brian Jones and Gerd Blanke

UDM (Unified Data Model) for chemical reactions – past, present and future

https://doi.org/10.1515/pac-2021-3013

Abstract: The UDM (Unified Data Model) is an open, extendable and freely available data format for the exchange of experimental information about compound synthesis and testing. The UDM had been initially

- *BioChemUDM – a unified data model for compounds and assays*
  - In 2017, same-day data exchange, no more spreadsheets by email
  - In 2019, Nurix Therapeutics applies to CDD Vault and PipelinePilot™
  - In 2021, IDEAYA Biosciences applies to AWS and Harmony-ML™



Editorial

Michael A. Kappler*, Christopher T. Lowden and J. Chris Culberson

BioChemUDM: a unified data model for compounds and assays

https://doi.org/10.1515/pac-2021-1004

Abstract: We present a simple, biochemistry data model (BioChemUDM) to represent compounds and assays for the purpose of capturing, reporting, and sharing data, both biological and chemical. We describe an

- *BiologyUDM – a unified data model for proteins and cell lines*

IDEAYA BIOSCIENCES

Structure as a Keystone

A Molecule-Centric View

- Compound identification is based purely on the chemical connection table with enhanced stereochemistry, so a compound identifier is based solely on a sketch
  - One assumption about stereocenters - unspecified means mixture
- There is no need for complicated business rules or controlled vocabularies to distinguish a compound
- Interpretation of sketches lends itself well to the language of the chemist (skilled in drawing molecules)

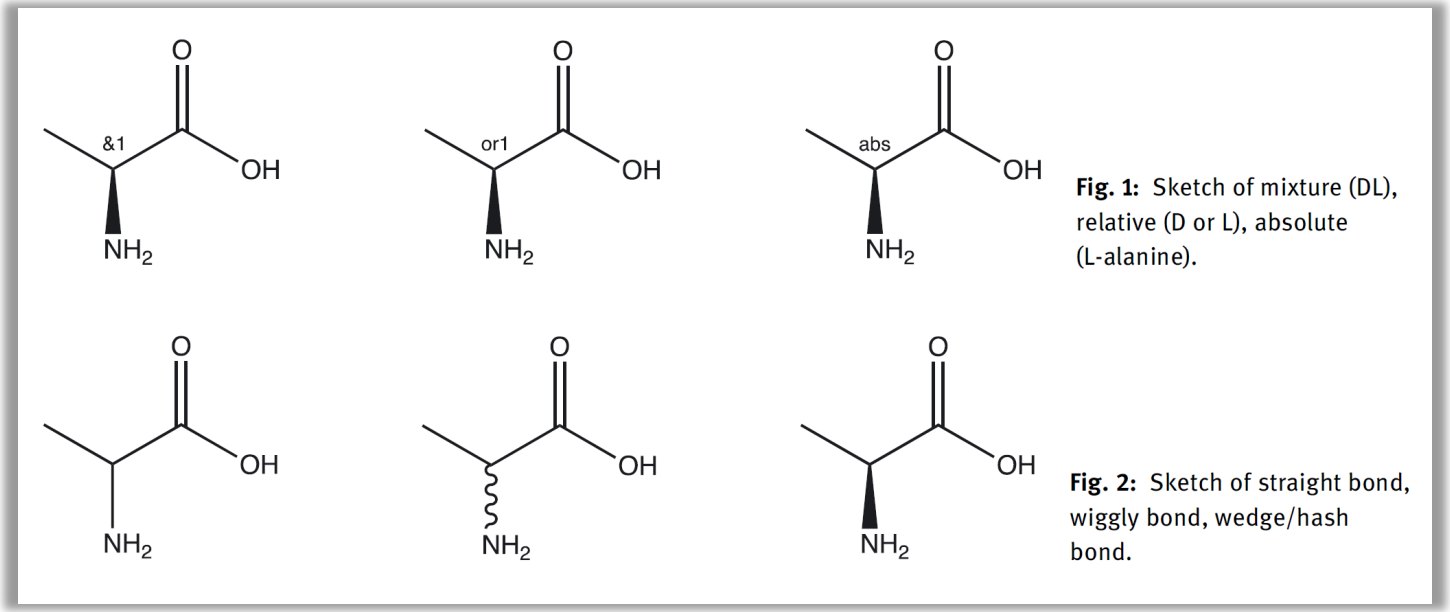


Fig. 1: Sketch of mixture (DL), relative (D or L), absolute (L-alanine).

Fig. 2: Sketch of straight bond, wiggly bond, wedge/hash bond.

How do I specify the structure?

Use one of these fields:

- UDM.MOL.Compound
- UDM.MOL.CXsmiles

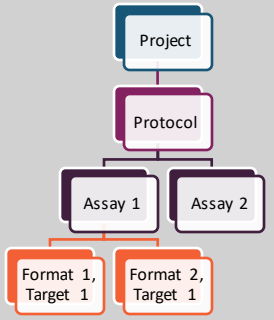
Hundreds of Assays in a Handful of Protocols

Accumulating Experiments is Easy

- Protocol categories are beneficial because similar assays can be treated in the same way.
- Within each category are fields to distinguish one assay from another.
  - For example, two kinds of activation are distinguished by the conditional field called 'Target' (AhR, PXR).
- To simplify the problem, we represent unpivoted data in categories:

Published
• Activation • Binding • Induction • Inhibition • Oxidation • Permeability • Pharmacokinetics • PhysChemProperty • Stability • Toxicity

Unpublished
• DoseResponseDn • DoseResponseUp • SurfacePR



How do I specify a molecule, batch, or sample as a test article?

Use one of these values:

- UDM.MOL.ID
- UDM.MOL.BAT.ID
- UDM.MOL.BAT.SAM.ID

…as the test article.

- UDM.ASY.TST.Article

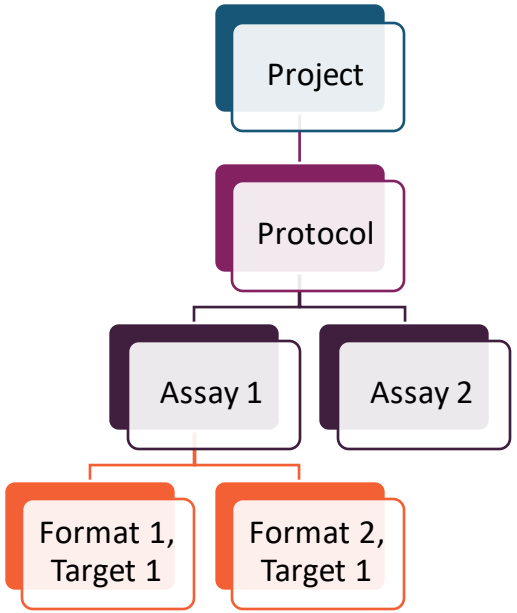
BioChemUDM: A Versatile Tool for Assay Capture

The Hierarchal Nature is Familiar and Easily Adopted by Data Generators

Hierarchal Model

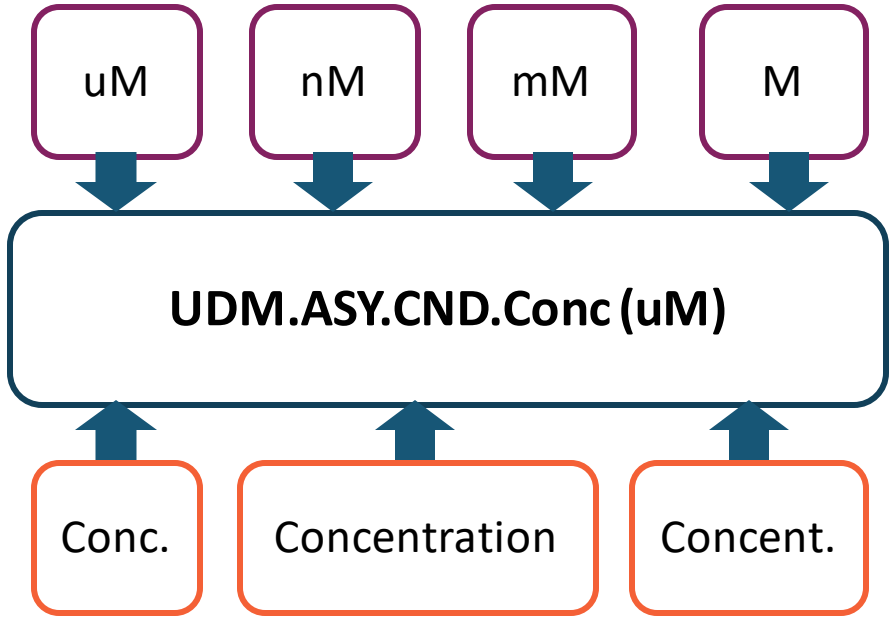
Scientists already have a fundamental understanding of the concepts and relationships used in the model

- Scientists understanding makes adoption obtainable



Data Standardization

- Set fields are straightforward and relevant to scientists and CROs



BioChemUDM: As We Grow

Adaptability and Scalability are Key Features

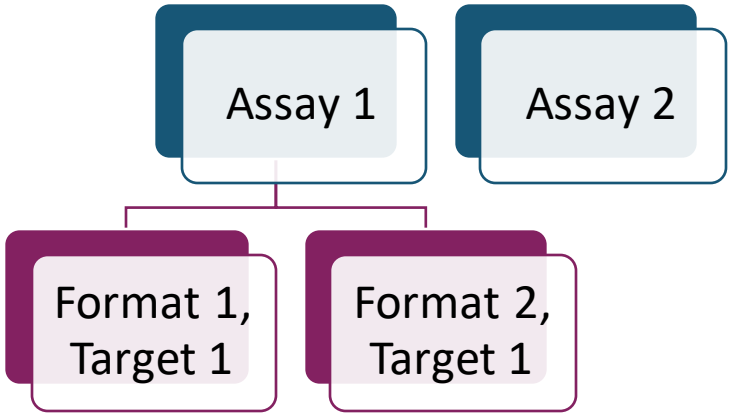
Adaptability

Describing a new assay is accomplished by:

- Utilizing existing conditional fields
  - Example: Use **UDM.ASY.CND.Conc (uM)** to define the concentration of the additive/stimulant based on the format of assay
- Extending the controlled vocabularies
  - Example: Use **UDM.ASY.CND.Target** to define the cell line or protein construct
- Creating an additional field
  - Example: Use a new field such as **UDM.ASY.RES.File** to enable capture of binary data

Scalability

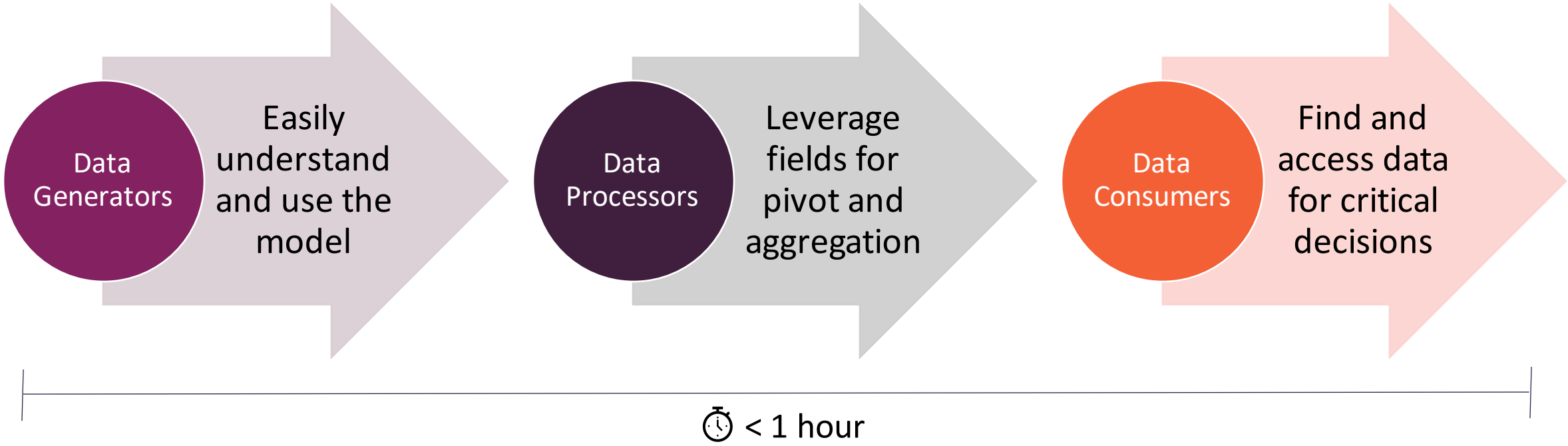
This entails the creation of new tags that describe a unique kind of assay



- Key Take Away: The component quality of the BioChemUDM means that each value within key fields can expand the amount of data captured

BioChemUDM: Racing Towards Answers

The Advantages Lead to Near Real-Time Data-Driven Decisions

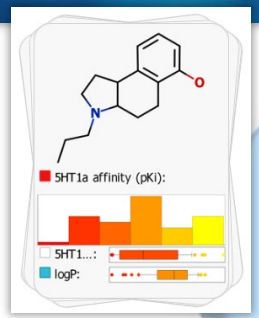


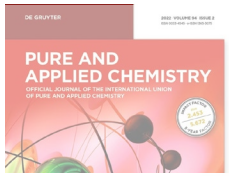
Modern Architecture to Broadly Leverage Information Across Domains

Platform & Visualization Landscape

optibrium™

StarDrop™





DE GRUYTER Pure Appl. Chem. 2022; aop

Editorial

Michael A. Kappler*, Christopher T. Lowden and J. Chris Culberson

BioChemUDM: a unified data model for compounds and assays

https://doi.org/10.1515/pac-2021-1004

Abstract: We present a simple, biochemistry data model (BioChemUDM) to represent compounds and assays for the purpose of capturing, reporting, and sharing data, both biological and chemical. We describe an approach to register a compound based solely on a stereo-enhanced sketch, thereby replacing the need for additional user-specified "flags" at the time of compound registration. We describe a convention for string-based labels that enables inter-organizational compound and assay data sharing. By co-adopting the Bio-ChemUDM, we have successfully enabled same-day exchange and utilization of chemical and biological information with various stakeholders.

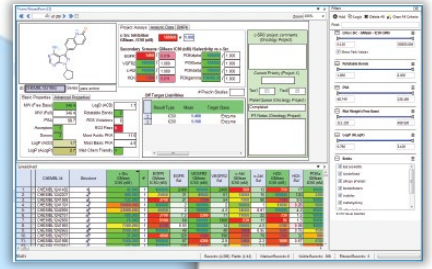
Keywords: Cheminformatics; data sharing; pharmaceutical informatics; unified data model.

SCHRÖDINGER.

LiveDesign®

CERTARA

D360



Data Lake

CDD VAULT®

mosaic

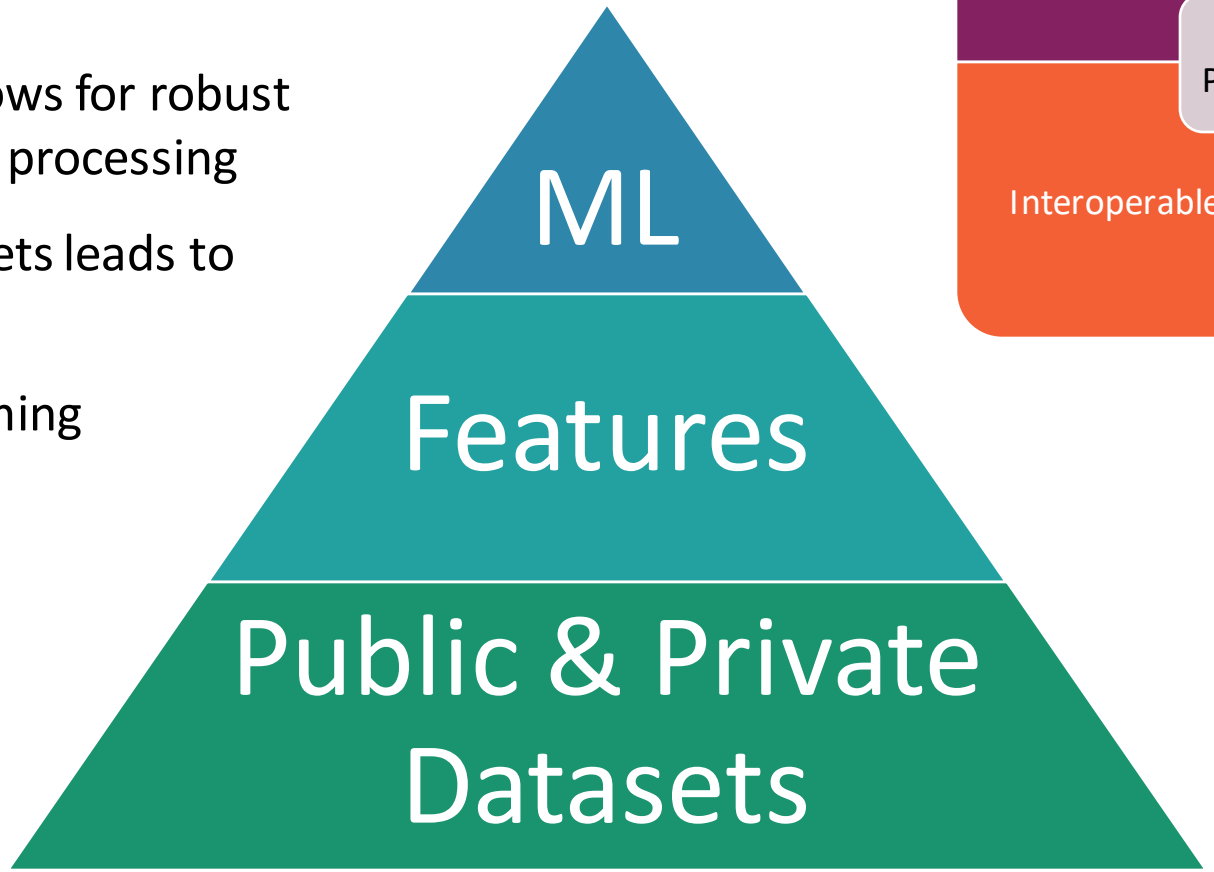
titian

IDEAYA BIOSCIENCES

Integrating the BioChemUDM with AWS

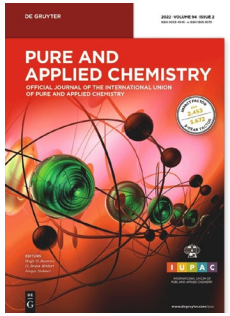
Trivializing Data Processing for Machine Learning

- Harmony-ML™, our machine learning factory, embraces FAIR principles
- Consistency in data capture allows for robust automation of undifferentiated processing
- Anticipated expansion of datasets leads to automated feature extraction
- The feature store leverages naming conventions and that's FAIR!



Machine-Learning Framework

Factory Concept based on AWS SageMaker



DE GRUYTER Pure Appl. Chem. 2022; aop

Editorial

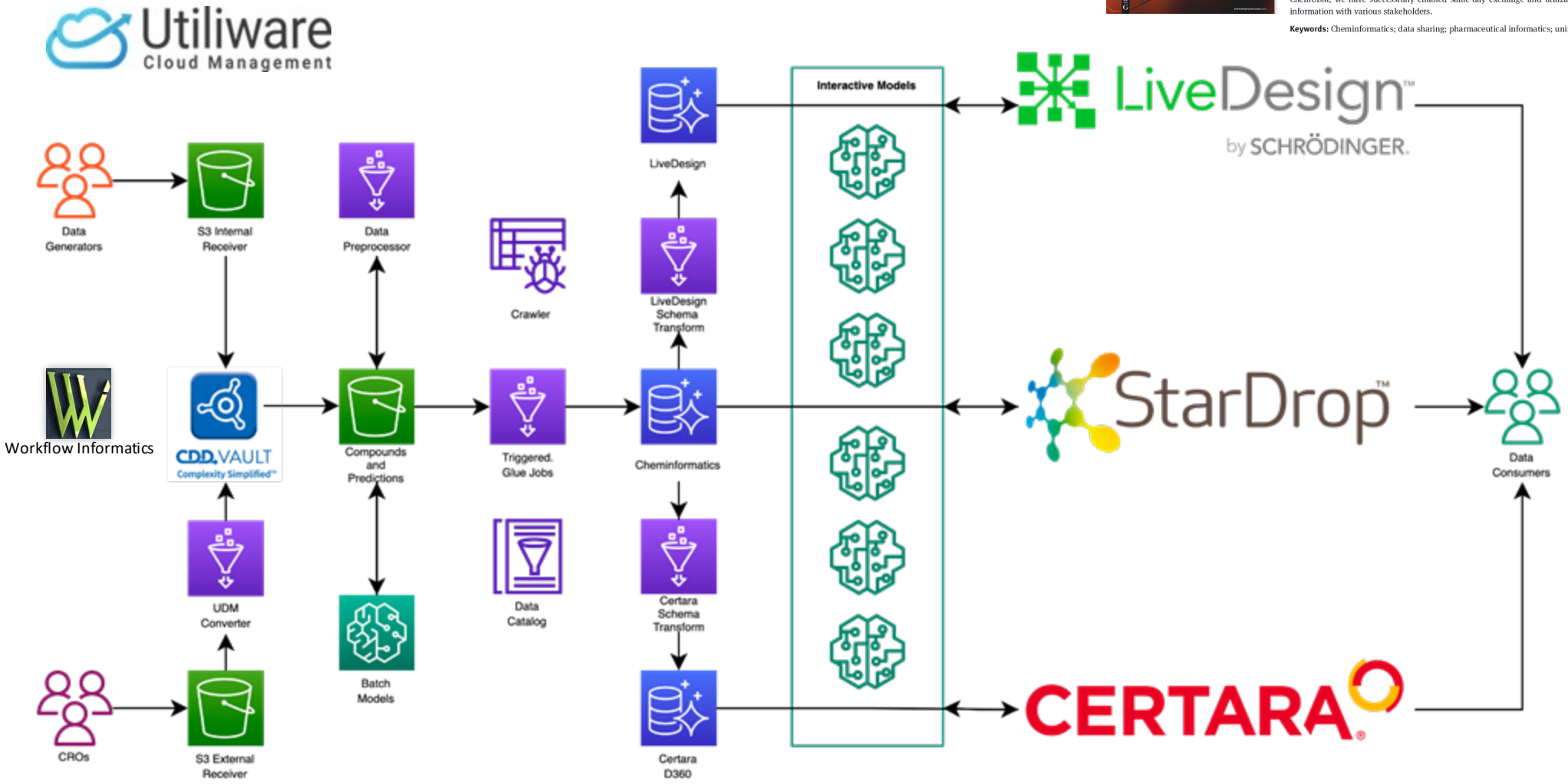
Michael A. Kappler*, Christopher T. Lowden and J. Chris Culberson

BioChemUDM: a unified data model for compounds and assays

https://doi.org/10.1515/pac-2021-1004

Abstract: We present a simple, biochemistry data model (BioChemUDM) to represent compounds and assays for the purpose of capturing, reporting, and sharing data, both biological and chemical. We describe an approach to register a compound based solely on a stereo-enhanced sketch, thereby replacing the need for additional user-specified "flags" at the time of compound registration. We describe a convention for string-based labels that enables inter-organizational compound and assay data sharing. By co-adopting the Bio-ChemUDM, we have successfully enabled same-day exchange and utilization of chemical and biological information with various stakeholders.

Keywords: Cheminformatics; data sharing; pharmaceutical informatics; unified data model.

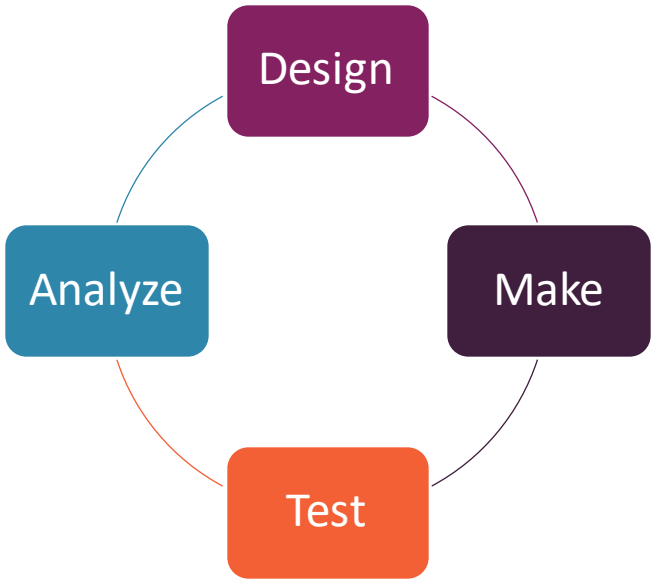


IDEAYA BIOSCIENCES

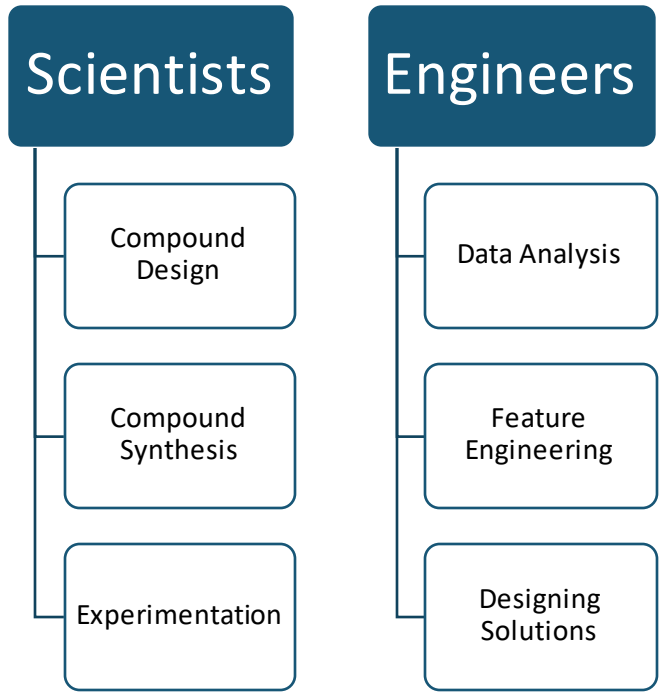
Harmony-ML™ Enhances the Drug Discovery Cycle

Making Data-Driven Decisions Faster; Freeing Time for Innovation

- Assay predictions improve compound prioritization
  - Triage ideas for compound progression
- Assay predictions lead to efficient cycle times
  - Scientists make better data-driven decisions



DOI:10.1016/j.drudis.2011.09.01

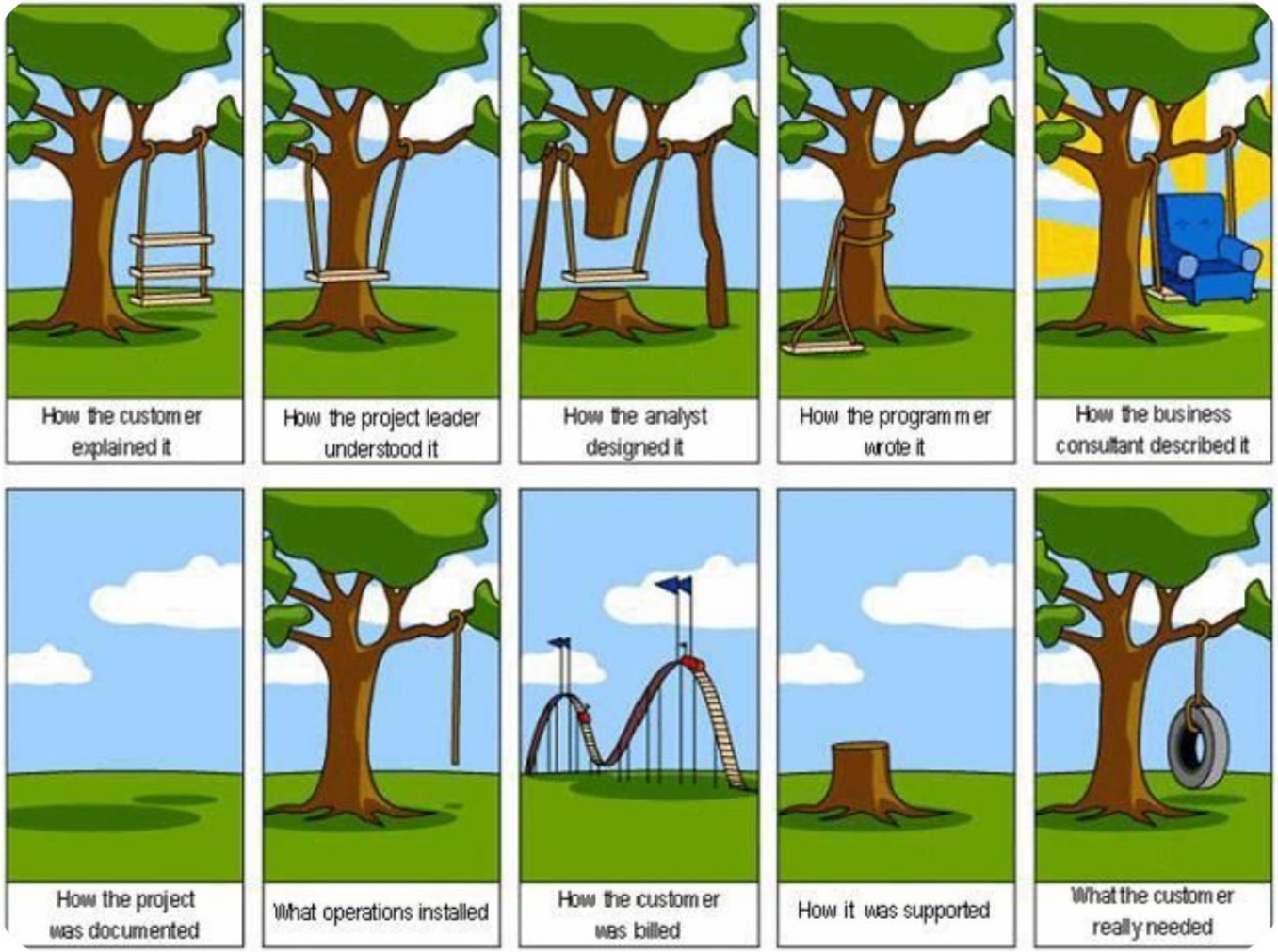


Summary and Acknowledgements

- We have described an approach to represent compounds and assay data from multiple, disparate sources using an extension of the Unified Data Model (BioChemUDM) concept
- This data model, like its predecessors, has been born out of necessity
- This work is the result of following the FAIR guiding principles
- Reducing compound identity to its connection table has benefits and consequences
- Scaling and adapting to user needs is a FAIRly simple process
- Through co-adoption, we can receive and share data with other research groups within the same day
- The BioChemUDM is a useful precursor for machine learning applications
- It is our hope the BioChemUDM, like the RxnUDM, will be embraced and incorporated into the UDM
- Plans for expansion to protein constructs and cell lines is underway
- Special Thanks: Utiliware, CDD, and Workflow Informatics for implementation support

What is an Informatics Platform?

It Depends on Your Perspective…



February 2023

Automating Drug Discovery Informatics

Talking Points on BioChemUDM, Software Selection/Implementation, Data Lake Tahoe, and Harmony-ML™