

Mass Spec Data Digital Transformation at Eli Lilly Biologics Discovery

2026 CASSS Mass Spec

Sep 16th, 2026

Joomi Ahn, PhD



Reinventing Drug Discovery with AI

Vision: Transition drug discovery from trial-and-error to simulation-driven engineering discipline

NVIDIA and Lilly Announce Co-Innovation AI Lab to Reinvent Drug Discovery In the Age of AI

January 12, 2026



Companies to Jointly Invest up to \$1 Billion Over Five Years in Infrastructure and Research

 [Download PDF](#)

News Summary:

- NVIDIA and Lilly bring together a world-leading, multidisciplinary team of scientists, AI researchers and engineers to address the hardest problems in drug discovery.
- The co-innovation lab infrastructure will be built on the NVIDIA BioNeMo platform and the NVIDIA Vera Rubin architecture.
- NVIDIA and Lilly will pioneer robotics and physical AI to accelerate and scale medicine discovery and production.

<https://investor.lilly.com/news-releases/news-release-details/nvidia-and-lilly-announce-co-innovation-ai-lab-reinvent-drug>

"Combining our volume of data and scientific knowledge with NVIDIA's computational power could reinvent drug discovery as we know it. We're creating the conditions for breakthroughs neither company could achieve alone."

— David A. Ricks, Chair & CEO, Lilly

"AI is transforming every industry, and its most profound impact will be in life sciences. NVIDIA and Lilly are bringing together the best of our industries to invent a new blueprint for drug discovery."

— Jensen Huang, Founder & CEO, NVIDIA

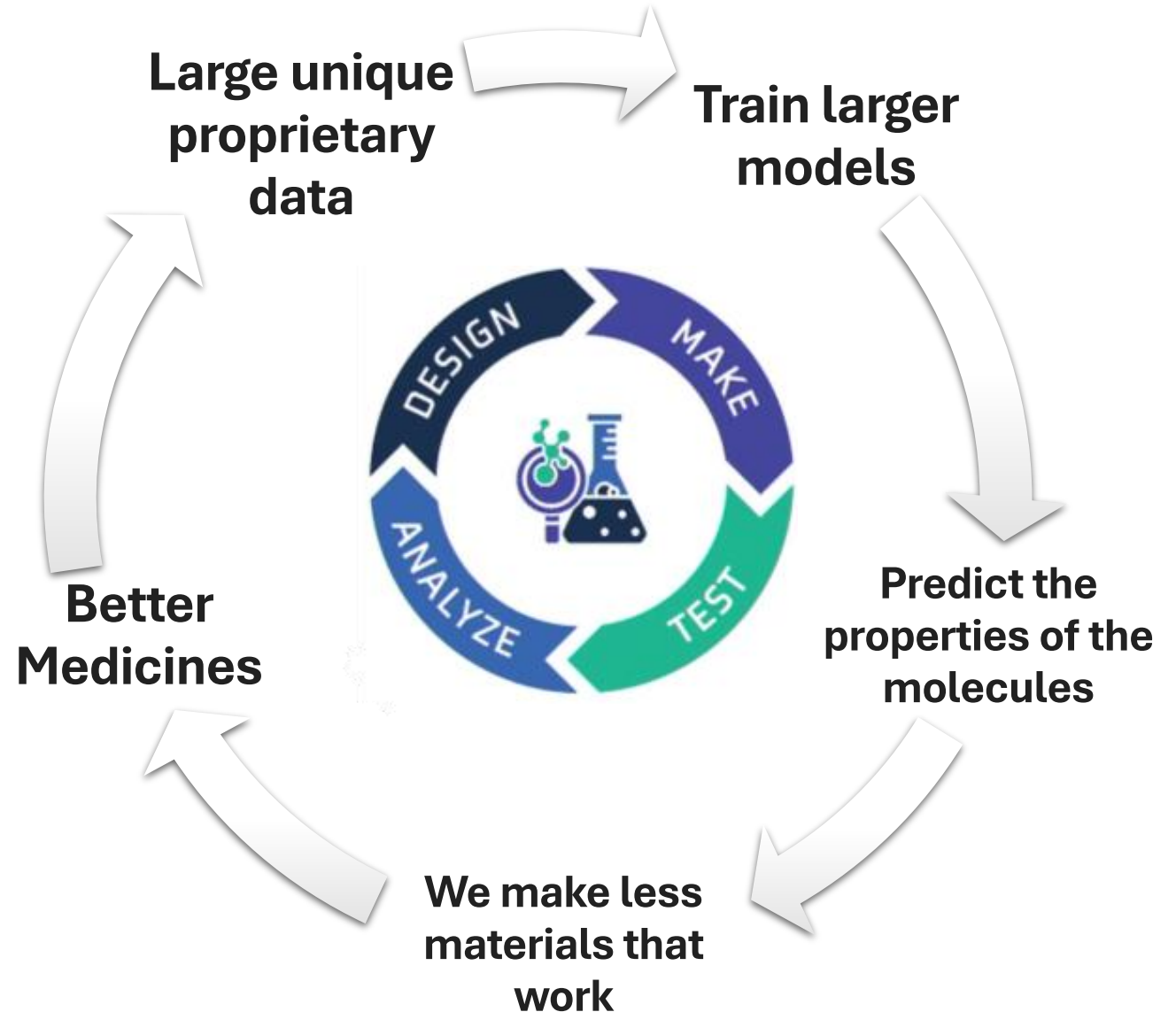
Faster Discovery Cycles and Better Model Performance Across the Pipeline

Our goals: Maximizing AI/ML capability by making a DMTA learning loop

- Algorithm
- Data
- Compute

Human Edge

- Critical thinking
- Scientific judgment



Modern Mass Spectrometry in Drug Discovery and Development in Biotherapeutics



Discovery

Many assays to assess antibody desirability:

- Epitopes
- Binding
- Functional assay
- Developability
- Stability
- Immunogenicity
- Clearance
- Solubility
- Viscosity
- Hydrophobicity
- Expression

Mass Spectrometry Applications

- ID
- Impurity
- Degradation hotspot
- Structural assembly
- Glycan
- MAM
- Many more...

Manufacturing

Development

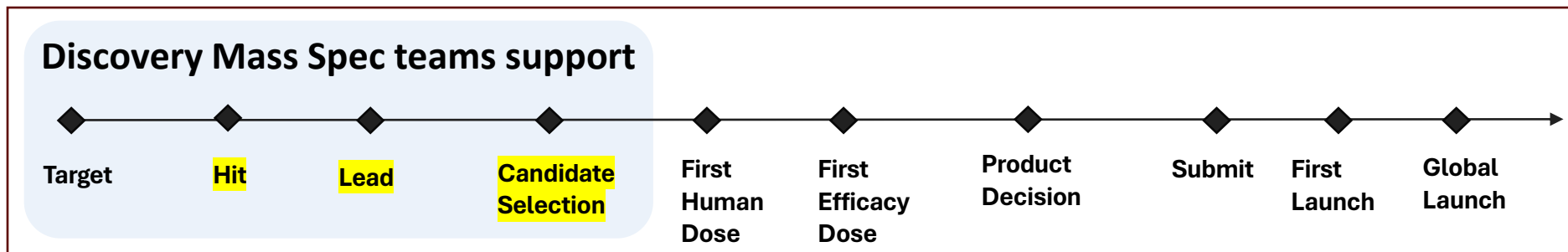
Discovery

We build MS Data Digital Transformation based on “FAIR” Data Principles

Findable	Accessible	Interoperable	Reusable
Searched by molecule and attribute, not by folder path	Retrieved by protocol under permission, not from a workstation	One question, one answer; across various vendor platforms	Acquired for one project, answering questions years later

- Enabling reuse by humans and machines (2016) and AI (2026)
- Contextual metadata tied to raw acquisitions
- Converts raw signals into structured results, captured as AI/ML-ready feature matrices
- Automated LIMS/eLN pipeline integration via enterprise platforms

What Was Our Roadblock in Mass Spec Wet-lab?

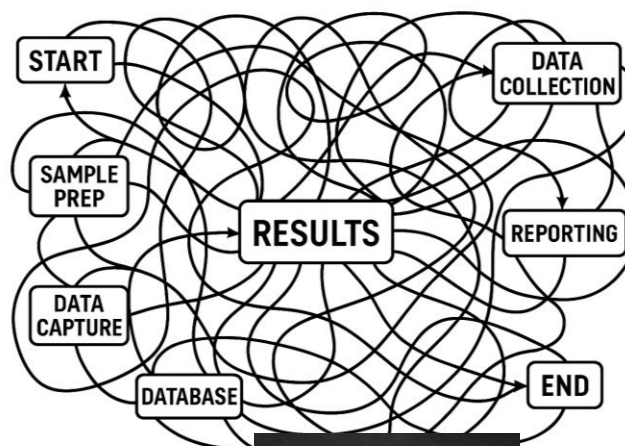


Our Mass Spec Tasks are Highly Complex

- Analyze various formats of molecules & modalities
- High throughput platform methods for screening purposes
- At the same time, low throughput methods for deeper characterization toward candidate selection (CS)
- Developability and stability assessment and impurity characterization



What is our roadblock?



Created a strategy roadmap

- Lilly BioTDR End-to-End platform
- Digital product landscape
- Internal products
- External products
- Align mass spec data digital strategy

What Our Mass Spec Scientists Need

- **Our needs:** Enterprise server-based vendor neutral MS data processing software / digital platform – we operate Thermo, Waters, Agilent, Bruker, and Sciex mass spec systems along with each vendor’s software and in-house software
- **Advantages**
 - Node-based workflow (no more black-box software)
 - Customizable reporting plugins
 - Comprehensive data review
 - Scalability in HT with automation
- **Three Step Deployment Process of the Server-based Software**
 - Local Integration Test
 - Development Server and User Acceptance Testing
 - Production Server Push

Building the Process by Connecting the Dots Around MS

1. HT Automated sample prep

2. LC/MS data acquisition

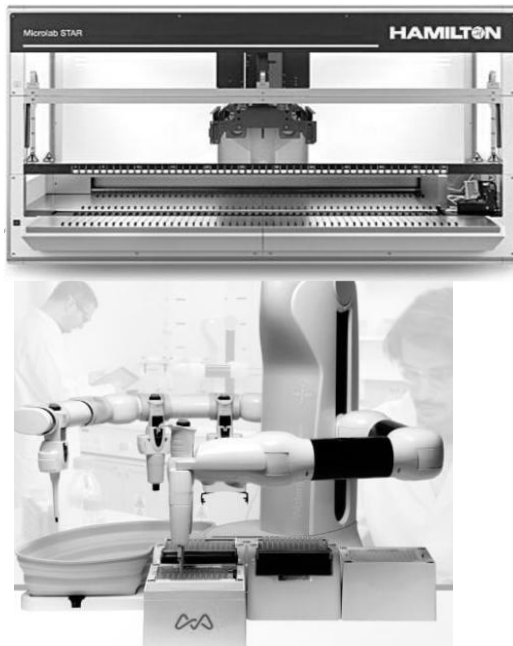
3. Automated raw file transfer using auto-push file watcher

4a. eLN notebook exports seq info in JSON & metafiles

4b. HT MS Data processing

5. Plug-in Customization & automated reporting

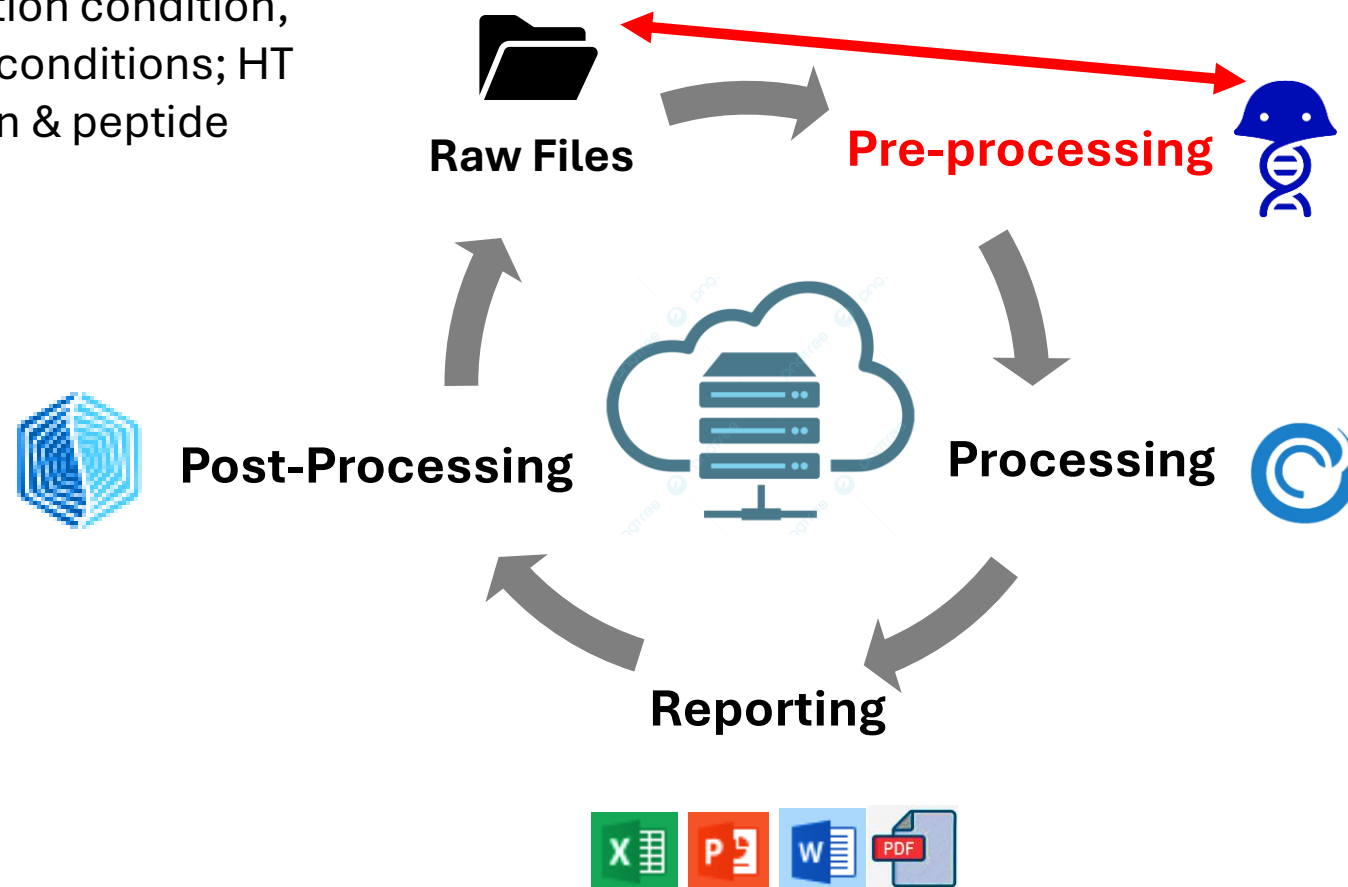
6. Build database & AI/ML prediction tools



How We Increase Efficiency using a Full Circle of Process:

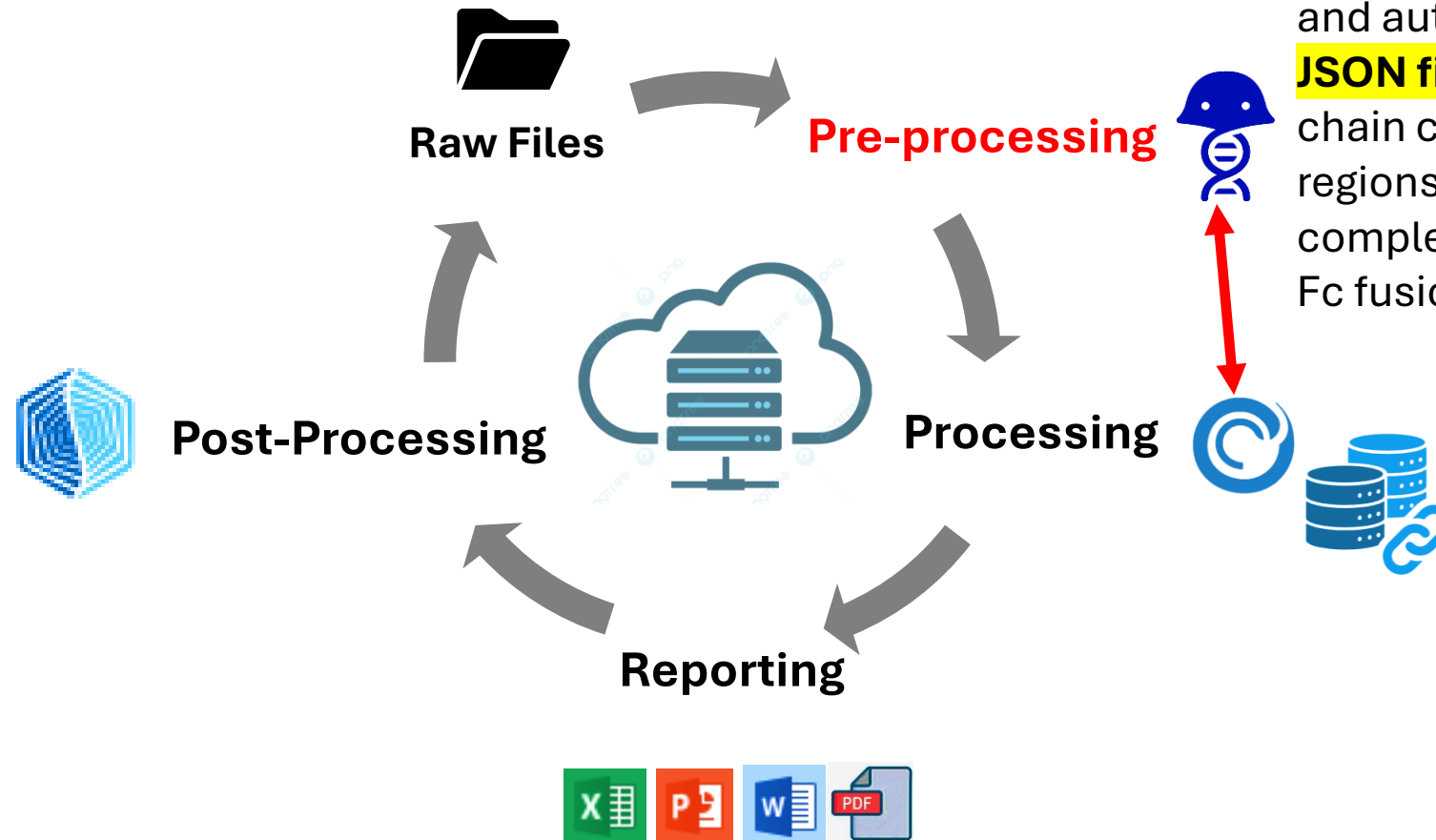
Pre-processing: No More Copy & Paste

Metafile generator automatically parse by molecule identifier, digestion condition, eLN notebook #, stability conditions; HT bispecific characterization & peptide mapping analyses



How We Increase Efficiency using a Full Circle of Process:

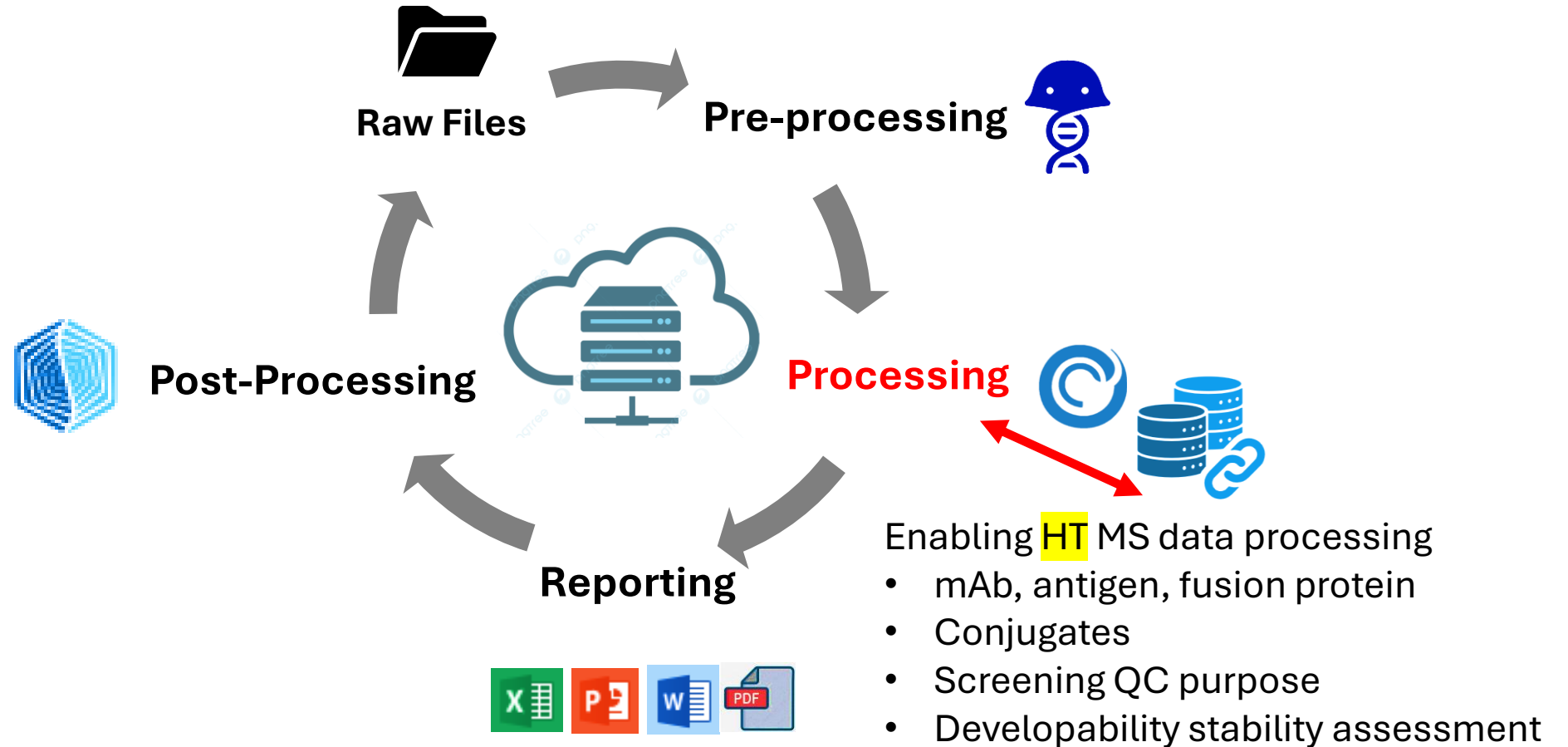
Pre-processing: No More Copy & Paste



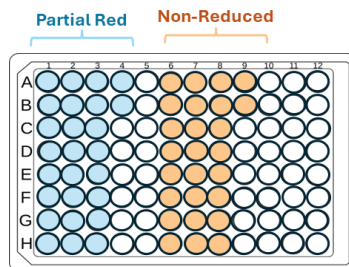
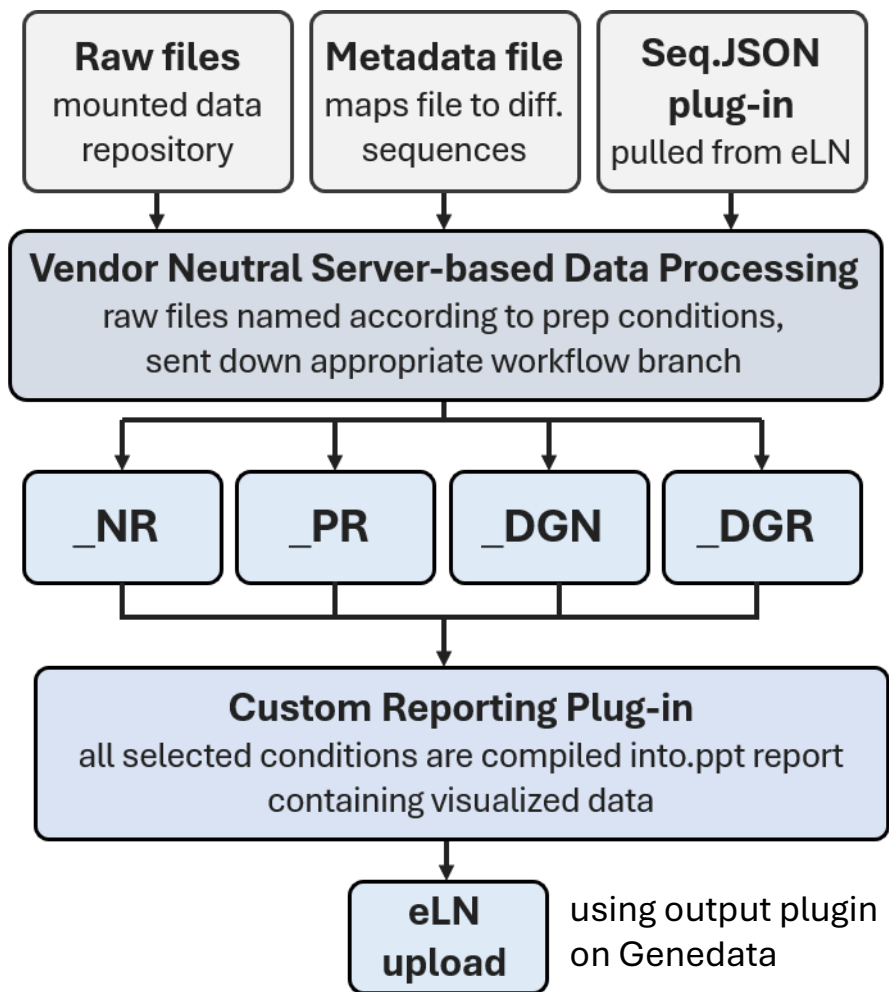
Benchling input plugin was built on Genedata expressionist to generate and automatically **pull JSON files into workflow** – chain configuration, CDR regions with sequences of complex modalities (scFv, Fc fusion, VHH, conjugates)

How We Increase Efficiency using a Full Circle of Process:

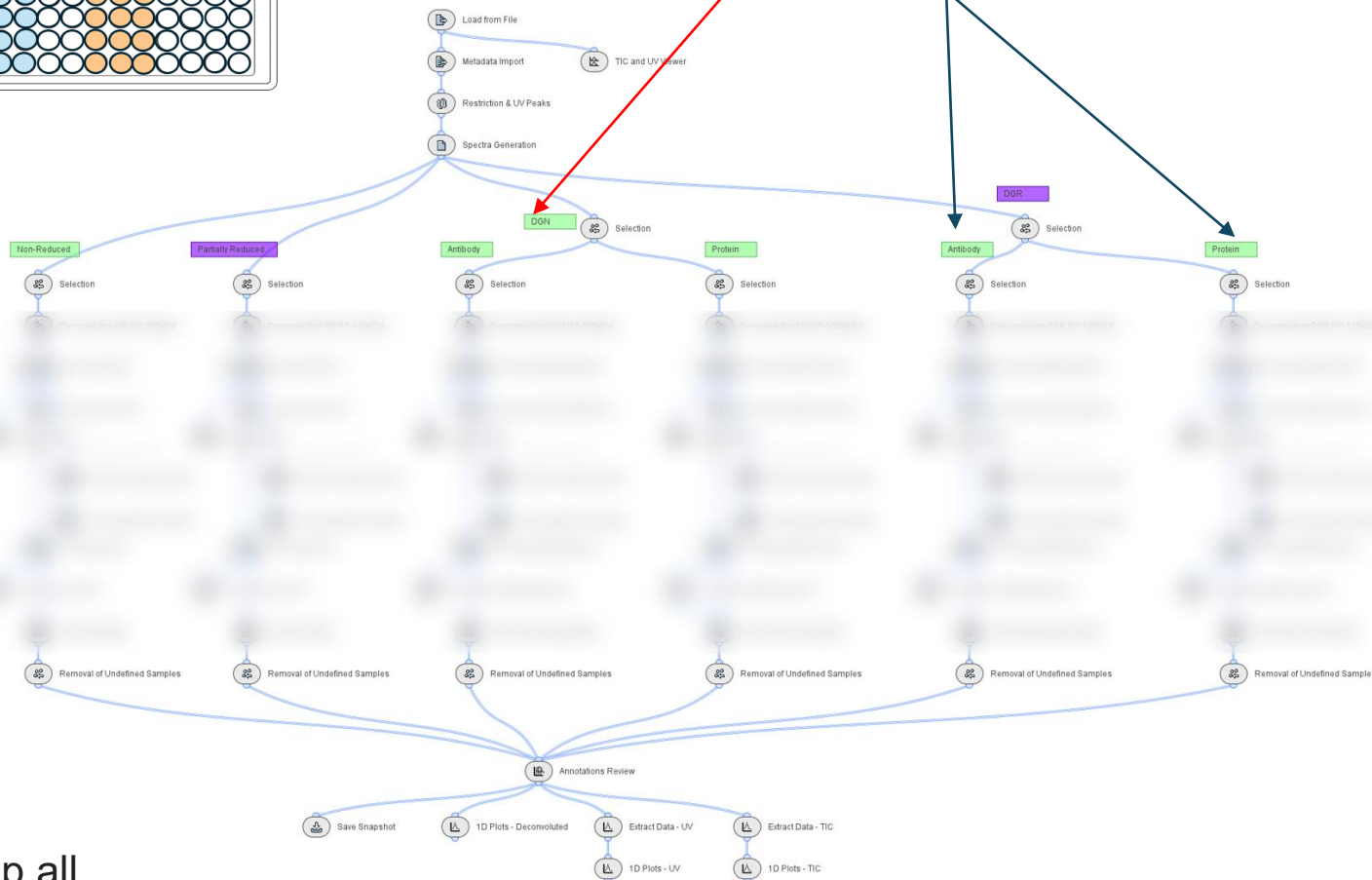
Processing



6 Examples of HT Intact / BsAb QC Workflows



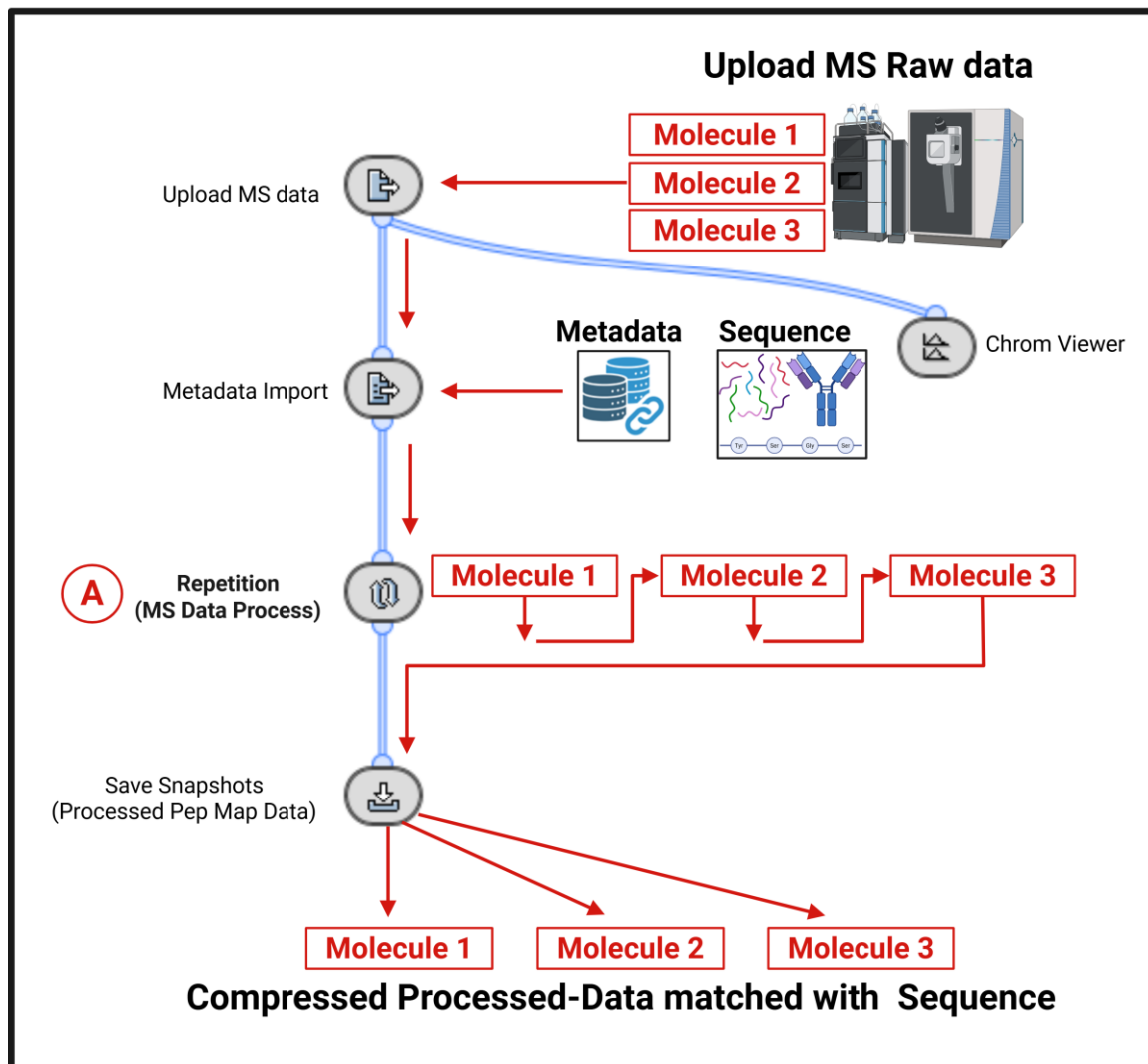
Data separated by
suffix (NR, Red, Degly, etc) and
molecule type



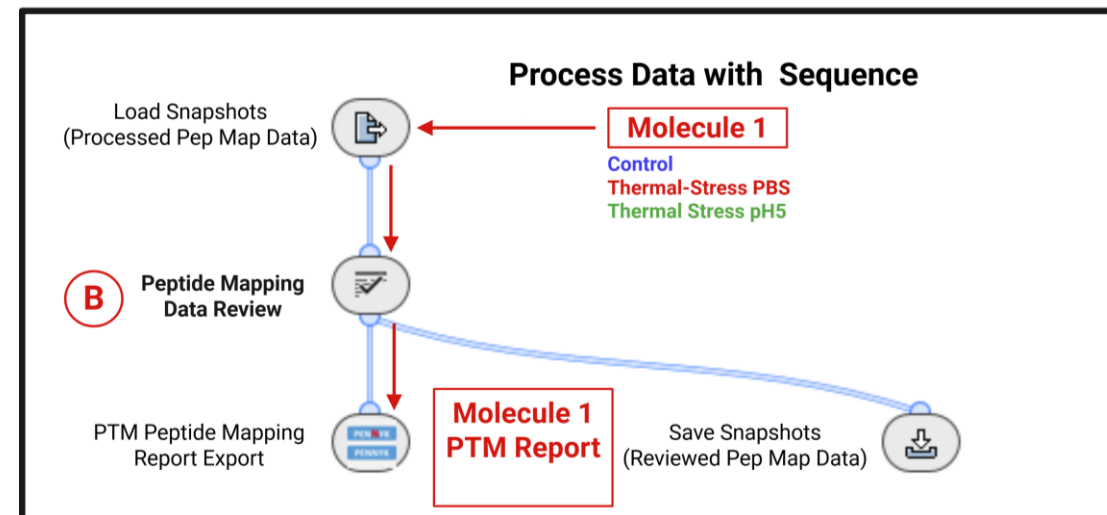
Modular, comprehensive workflow allows user to dump all kinds of sample set **AND** process with customized mass ranges, disulfide patterns, and glycosylation status searches.

Example of HT Peptide Mapping **Stability Workflow**

Data Process Workflow



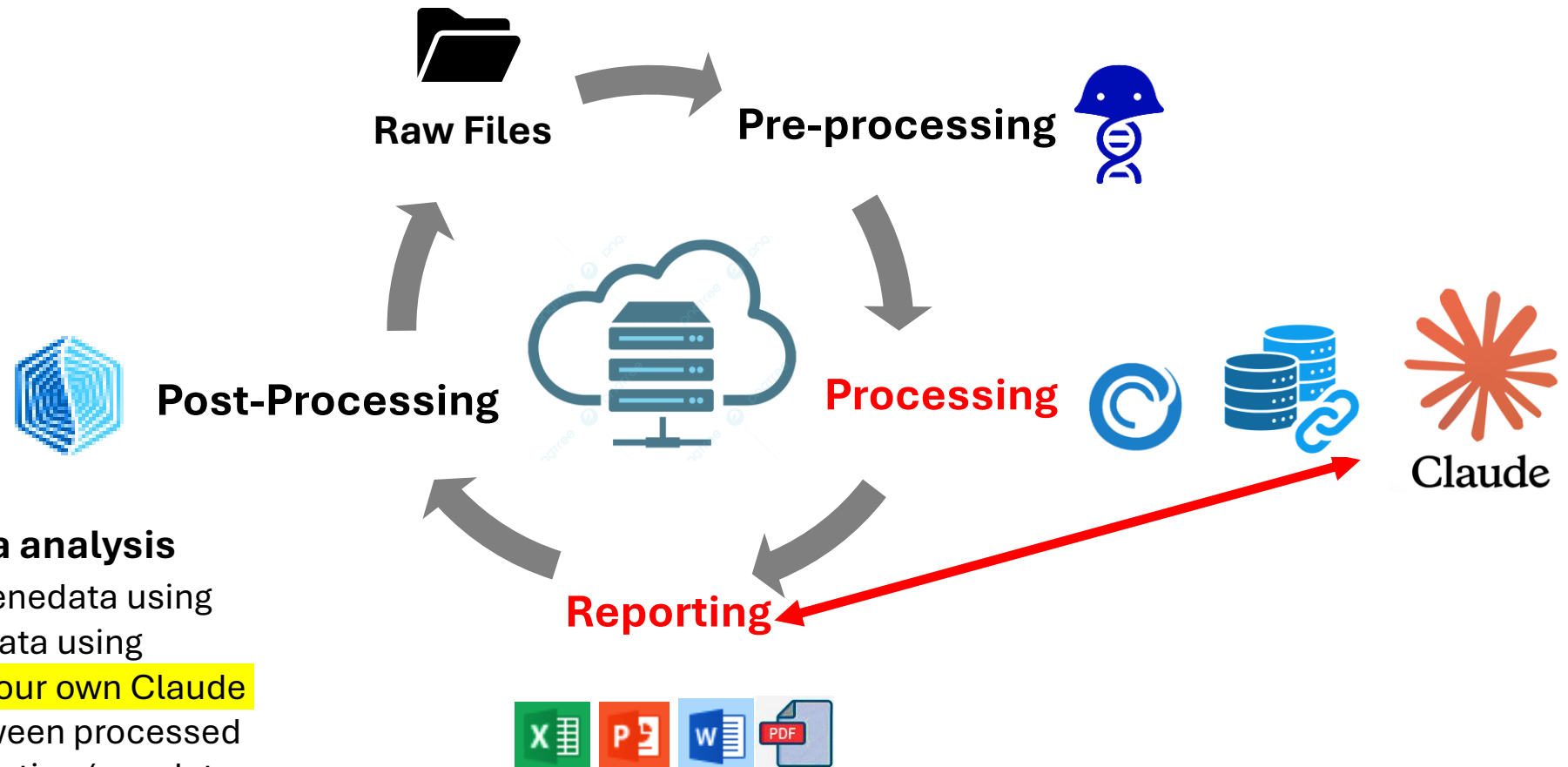
Data Review Workflow



- Repetition
- Automatic peptide review: ambiguous annotation, isomer detection, modification RT window set up, MSMS coverage,
- Improved visualization of fragment spectra annotations as a sum of all spectra → removed time-consuming steps

RESULT: Reduced turnaround time from weeks to hours. (Productivity increase at least 3 times)

Now MS Scientists Build Our Own Tools using Claude



Highly customizable data analysis tools were built inside of Genedata using plugins and outside of Genedata using Claude. **MS scientists create our own Claude tools to help fill the gaps** between processed data and structured data reporting (e.g. data organizer tools)

Post-processing Step in Full Circle

No More Excel tables, PowerPoint slides, PDF files

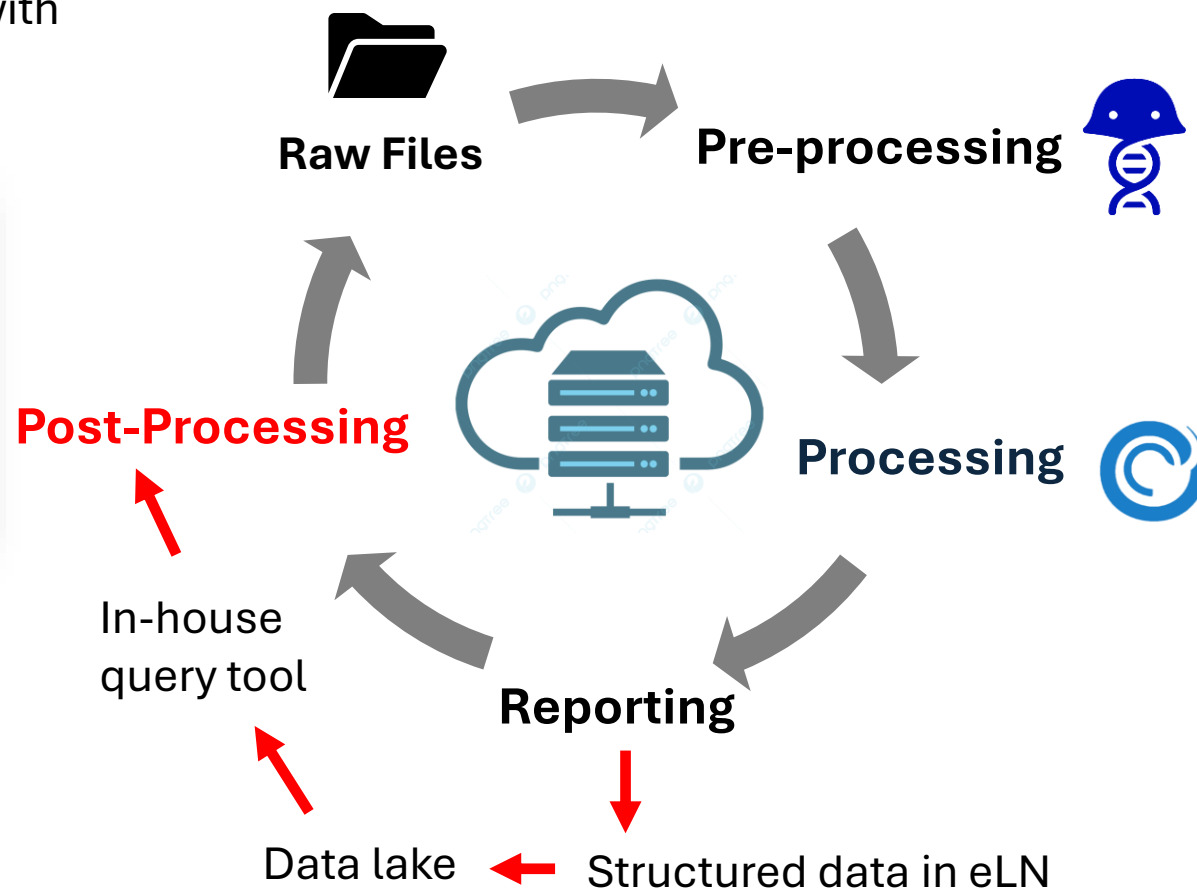
Stability mass spec data (intact & peptide mapping) are combined with SEC, cIEF, & CE-SDS

Markdown Report Generator

Generate comprehensive stability reports from research data in Markdown format

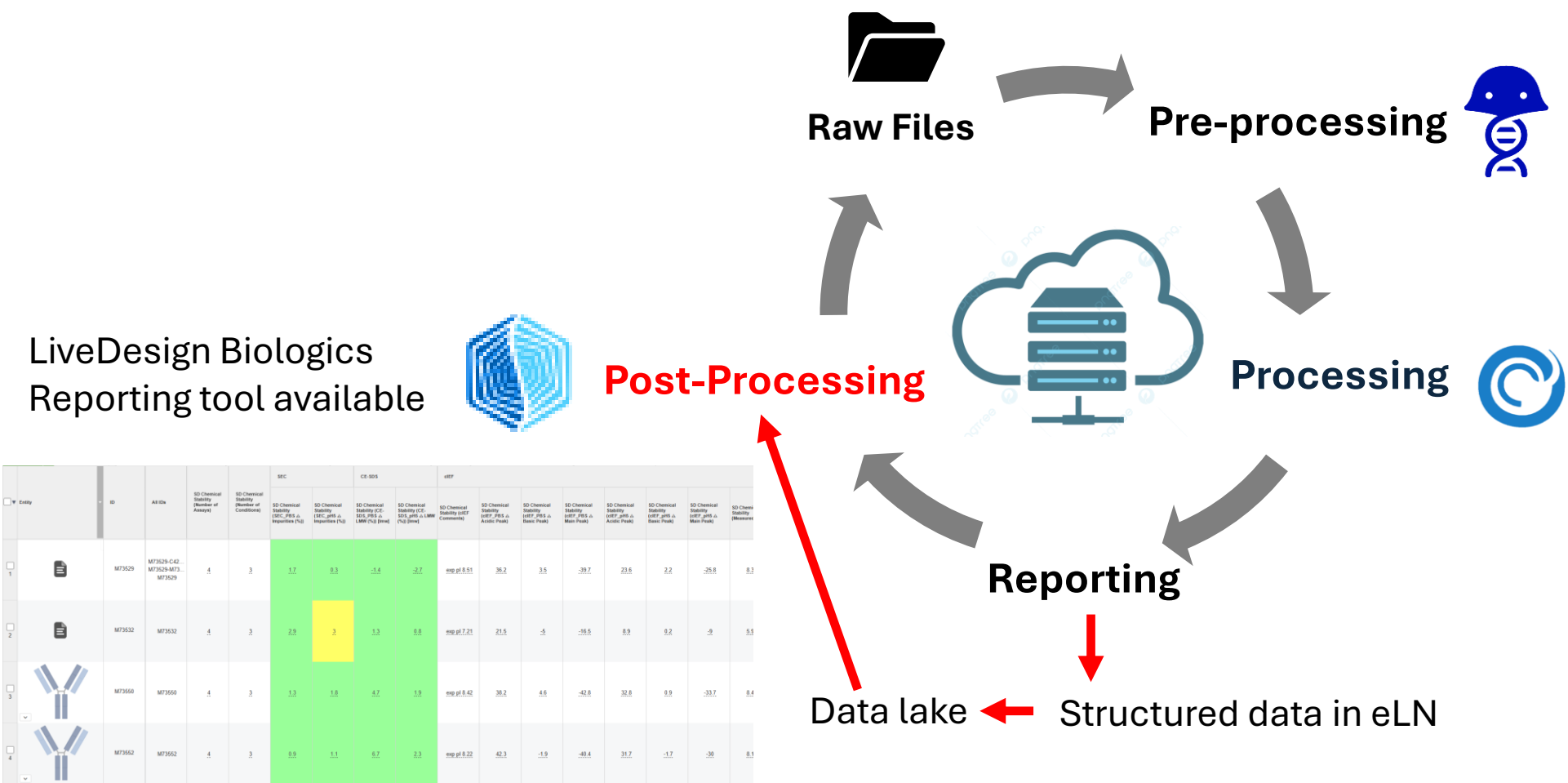
[Go to Report Generator](#)

HTML based markdown reporting tool automatically pulls all of stability results out of structured data capture from Benchling

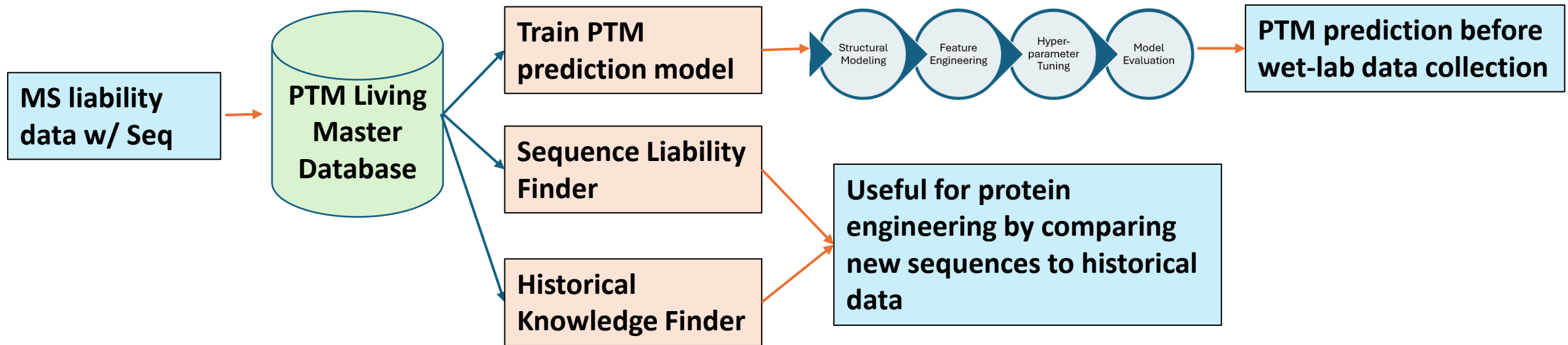


Report Automation Tool Beyond MS Data

Stability data package of SEC, cIEF, CE-SDS, intact mass, and peptide mapping

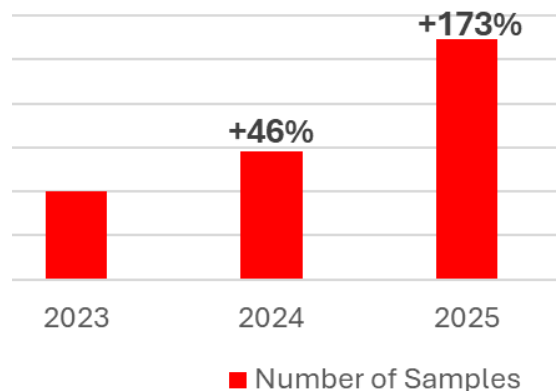


Living Database to Prediction Model Implementation



Values and Impact

Productivity Increase



- Given same time and resource, **overall productivity increase =173% over 3 years**
- Expect to further increase up to 10 times in next coming years

Data Integrity



From Unstructured Raw/Processed Data

- Every step in our process was manual
- Lots of copy & paste



To Structured Data

- Automated import of source to report
- Human and machine-readable output
- Data-driven raw to processed

Easier access to our knowledge



- Historical knowledge with specific experimental conditions is available in our hands
- Future: Product-specific AI feeding our data to Enterprise AI

Acknowledgments

- Shayan Akhavan
 - Aaron Cazy
 - Koyuki Takenaka
 - Olivia Huffman
 - Wen Chen
 - Andy Chau
 - Michael Batt
 - Amber M Peariso
 - Chris Kovach
 - Xianyin Lai
 - Di Wu
 - Yuhao Lin
 - Zach Alexander Rollins
 - Joomi Ahn
 - Ed Simpson
 - Peter Hillier
 - Angela Lee
 - Brendon Desmond
 - Xiangnan Dang
 - Melody Shahsavarian
 - Katie Brannon Corwin
 - Yue Webster Wang
 - Pavani Sudha Guttula
 - Jhansi Dharmaraj
 - Anshul Kanakia
 - John Stallings
 - Justin Walters
 - Bruno dos Santos
- Genedata Expressionist
- Stephen Kok
 - Daniel Abegg