

HUPO-PSI標準 小分子質量分析定量結果報告用データフォーマット

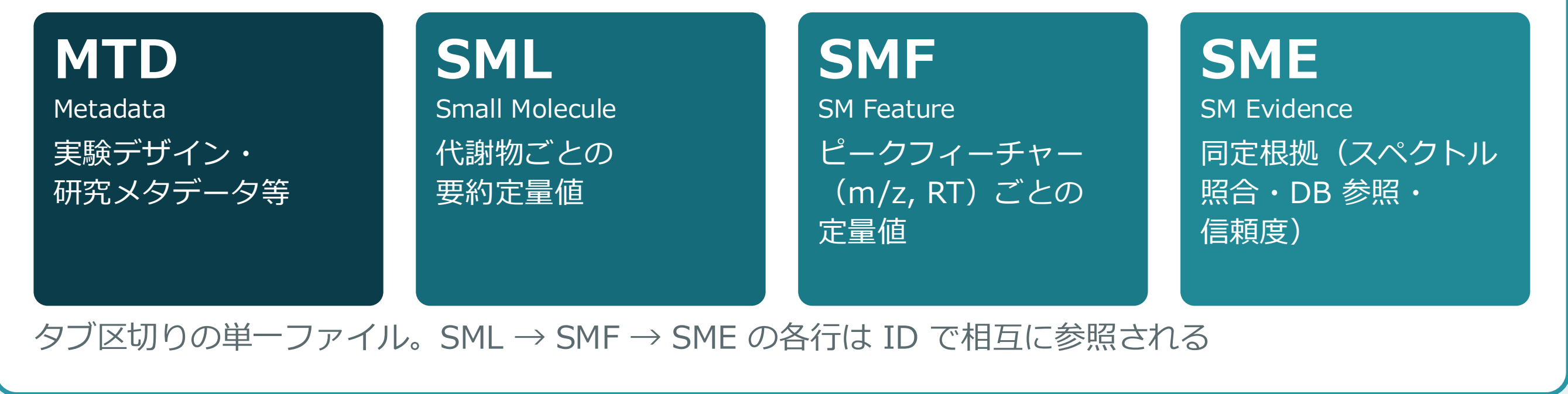
mzTab-M のAIを活用したデータ拡充とその利用

西田 孝三¹、Gabriele Tomè²、Johannes Rainer²、Ozgur Yurekten³、福島敦史⁴、津川裕司⁵、早川英介⁶、Philippine Louail²、Nils Hoffmann⁷

1) 理化学研究所、2) Eurac Research、3) EMBL-EBI、4) 京都府立大学、5) 京都大学、6) 九州工業大学、7) Forschungszentrum Jülich GmbH

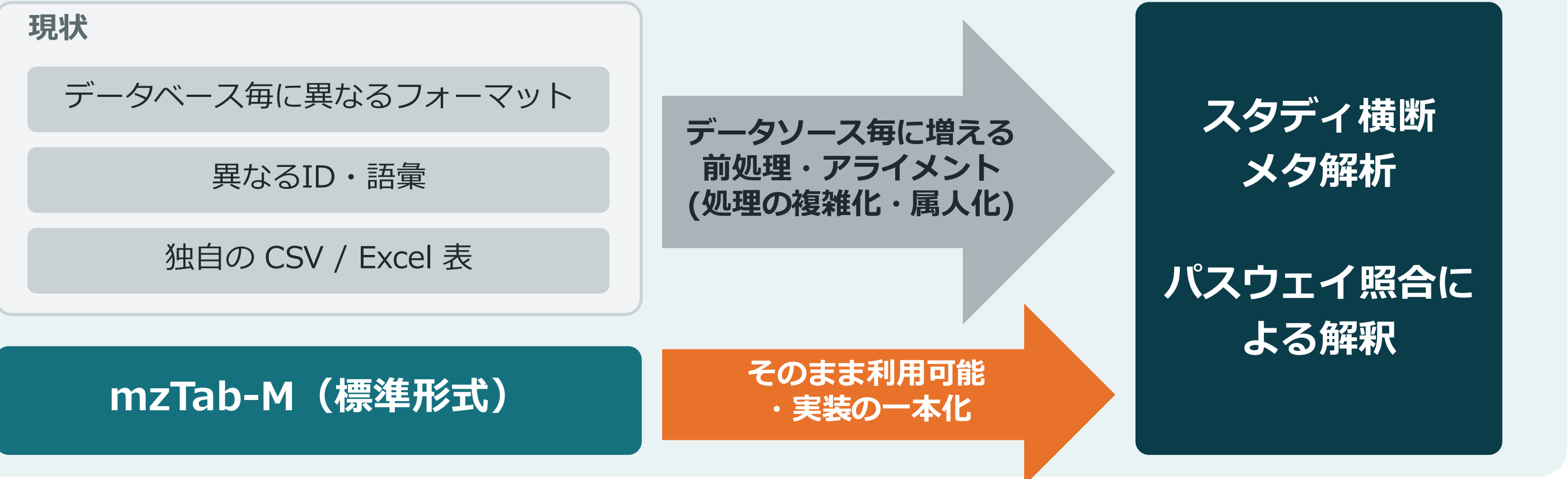
背景：mzTab-M とは

HUPO-PSI が策定した、小分子（メタボローム）質量分析の定量結果を報告するための標準データフォーマット。4 つのセクションをタブ区切りの単一ファイルに収め、定量値・同定根拠・実験デザインを機械可読な形で一体的に記述できる。厳格なスキーマを持ち、AI readyである。



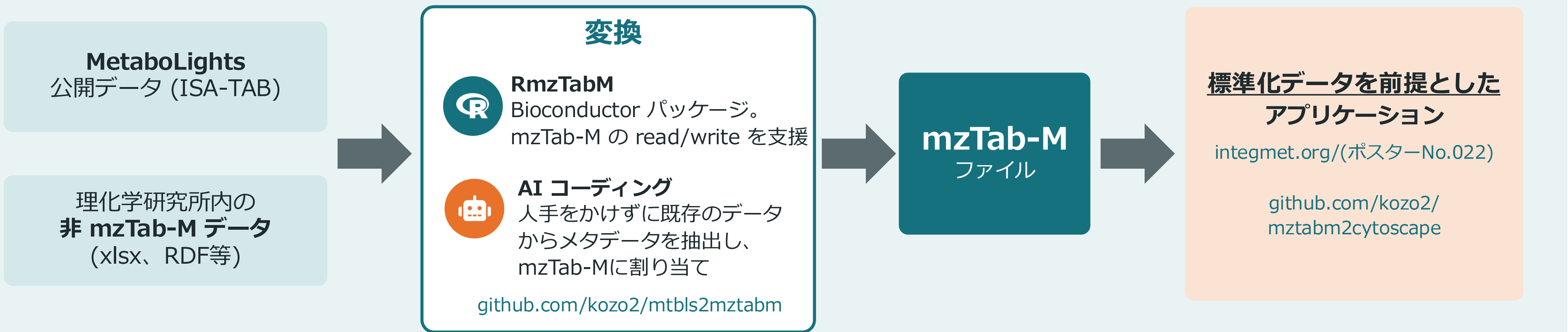
課題：公共データの多くは独自フォーマット

公共の定量後の質量分析メタボロミクスデータは依然として(mzTab-Mの論文出版は2019年)独自の形式であり、スタディ横断のメタ解析等は容易ではない。mzTab-M はこの障壁を取り除き、再利用性と相互運用性を高める鍵となる。



取り組み：AIを活用し、mzTab-M データを増やす・活用事例を広める

筆者は HUPO-PSI mzTab ワーキンググループの一員として、mzTab-M 形式のデータを増やすことと、その活用事例を広めることに取り組んでいる。これまでは既存の独自フォーマットのデータを mzTab-M 化することは容易ではなかったが、現在は RmzTabM パッケージと AI コーディング エージェントを併用することでデータ拡充が容易になり、その利活用例を作ることも可能になってきた。



AIが支援するmzTab-M化の手順 ([]はオプション)

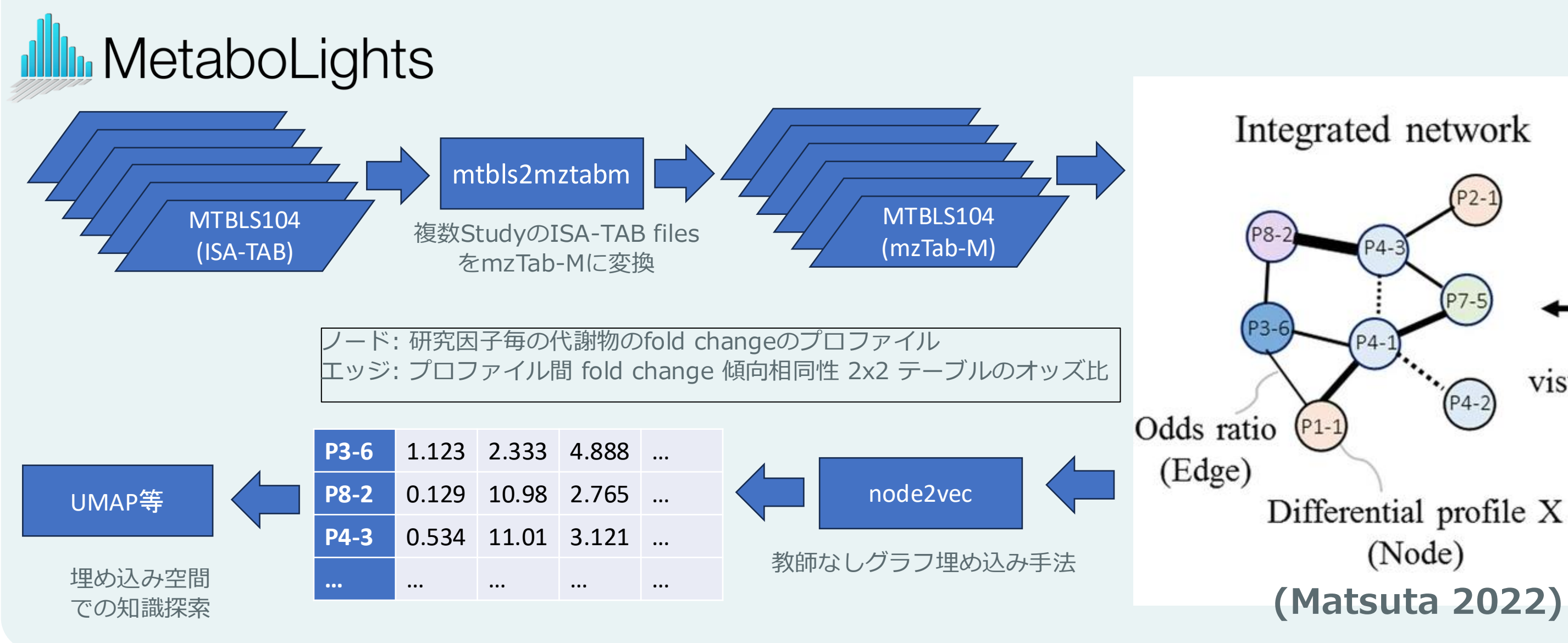
- 1 元データの把握 フィーチャーテーブル・同定結果・サンプル情報の所在と形式を確認する
- 2 MTD の記述 実験デザイン・アッセイ・研究変数・ソフトウェアを CV 用語で記述する
- 3 SML [/ SMF の生成] 定量値を代謝物とフィーチャーに対応付け、ID で紐付ける
- 4 [SME の記述] [同定根拠 (スペクトル照合・DB 参照・信頼度) を明示する]
- 5 検証 バリデータで構文・意味論を検査し、「人」が内容を確認する

mzTab-M データ作成における留意点

- ✓ AI エージェントが生成した変換コードや、CV 用語の選択は人が検証する必要あり
- ✓ メタボロミクス分野の質量分析手法は多様であるため、mzTab-M データ作成用コードも「単純明快」とはならない。「完成」も無い。
- ✓ 「既存」データでピークフィーチャーの再現性を担保、もしくは同定根拠をデータとして残しているものはほぼ無いため、SMF/SME までは書けないことがほとんど
- ✓ mzTab-M に対してメタデータの完璧性を求めると、データ作成が終わらない。アプリケーションを見据え、その必要条件のデータだけを含む mzTab-M でもよい
- ✓ mzTab-M は semantic validation (CV mapping) 機能を持つが、本件はその機能の活用までは至っていない

応用例① スタディ横断ネットワーク解析

MetaboLights データベースの異なるスタディ間で、代謝物の発現変動傾向の類似性ネットワークを構築(iDMET 手法 [Matsuta 2022])し、そのグラフエンベディングを解析する試み (integMET におけるネットワーク探索の試み)。mzTab-M により定量値と研究変数を統一的に取得できるため、メタ解析用コードの実装が明快になる。

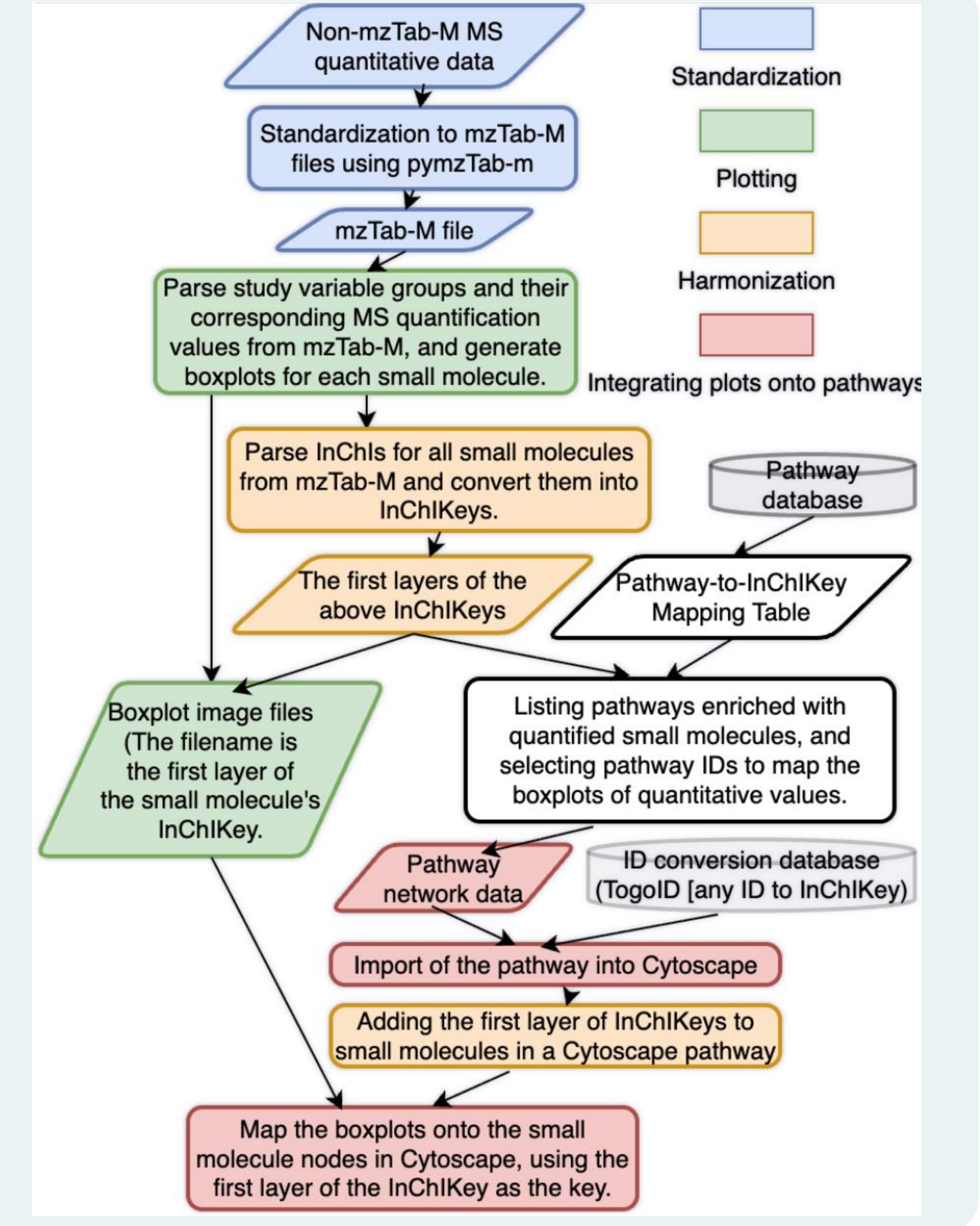


応用例② パスウェイ統合と Cytoscape 可視化

mzTab-M の同定結果と定量値を代謝パスウェイ情報と統合し、Cytoscape で可視化するパイプライン(mztabm2cytoscape)。

入力をmzTab-Mフォーマットに一本化しているため、煩雑な「データソース毎の前処理/データソース間の統合」が不要に。

これまで属人的に行われていた質量分析メタボロームのパスウェイ可視化にAI readyなデータフローを与えることが可能に。



まとめ

mzTab-M を共通基盤とすることで、「個別スタディに閉じない横断的な解析」や「これまで属人的に行われていたワークフロー」がAIで高速に実現可能になる。変換の実務知見と活用事例を共有し、mzTab-M 形式のデータとアプリケーションの両輪でエコシステムを広げていく。



このポスターは クリエイティブ・コモンズ 表示 4.0 国際 ライセンス (CC BY 4.0) の下に提供されています。
<https://creativecommons.org/licenses/by/4.0/deed.ja>

mzTab-M 仕様 <https://github.com/HUPO-PSI/mzTab>
RmzTabM <https://github.com/rformassspectrometry/RmzTabM>
MetaboLights ISA-TAB から mzTab-M への変換 <https://github.com/kozo2/mtbls2mztabm>
mzTab-M とパスウェイの統合可視化の自動化 <https://github.com/kozo2/mztabm2cytoscape>