

これが「構造」

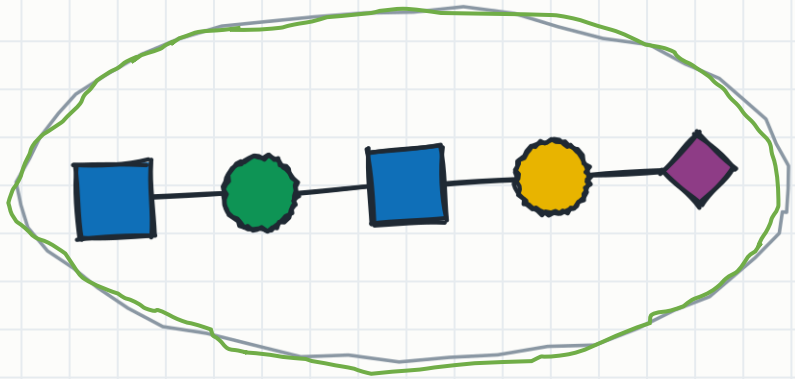
そのデータ、
AI は何だと思って読むのか

実験の文脈を残すための、メタデータ基盤の話

木下 聖子（創価大学）

トーゴーの日シンポジウム 2026 ／ 2026年10月5日

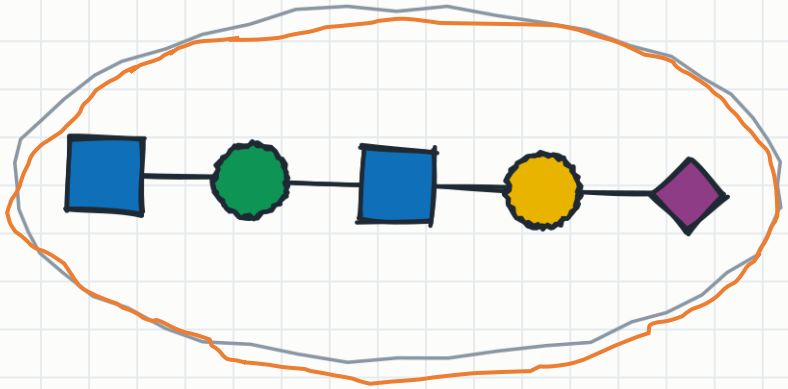
同じ構造でも、意味は同じではない



組織 A ・ 正常

発現量 / 局在 / 修飾の文脈





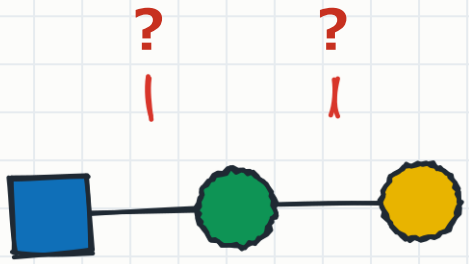
組織 B ・ 疾患

同じ構造記述、異なる意味

構造だけでは、生物学的な意味は決まらない。

由来組織・疾患状態・測定法が落ちた瞬間、その構造は「ただの文字列」になる。

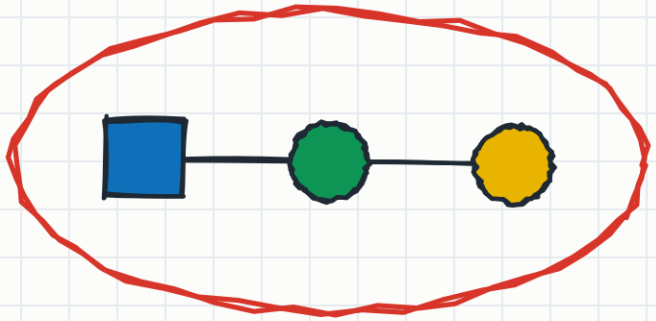
「測っていない」ことは、記録に残らない



実験で分かったこと

結合位置・アノマー配置は
手法によっては決まらない





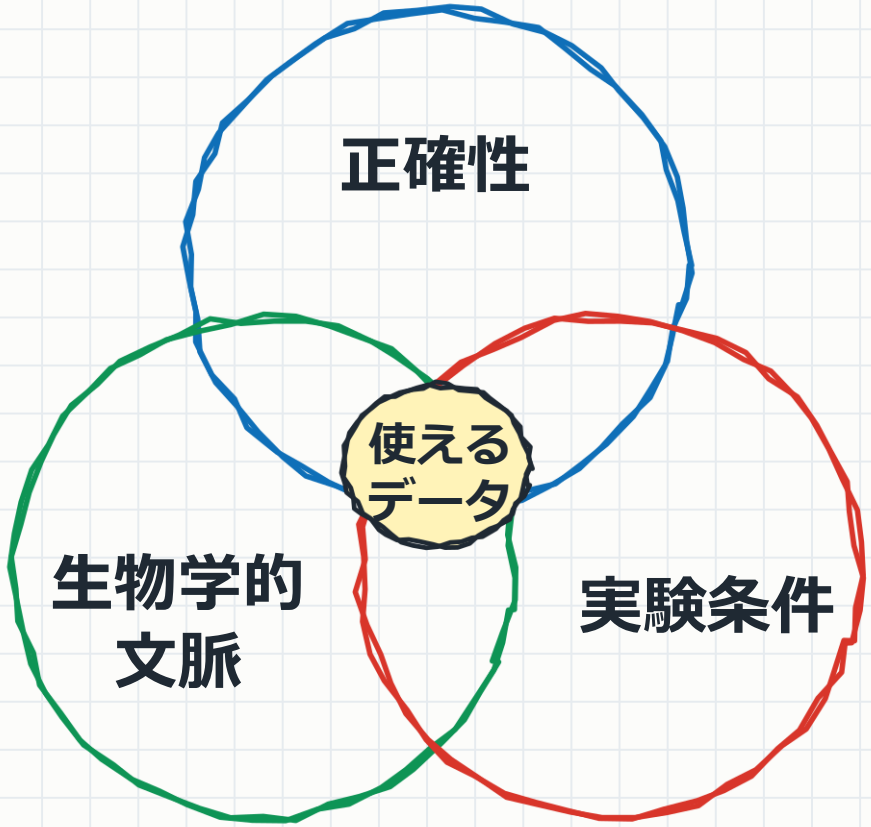
AI が読み取ること

確定した構造として扱われる
不確かさは伝わらない

欠けている情報ほど記録に残らず、そのまま「無かったこと」になる。

人間の読み手は行間で補える。学習するモデルは補わない。

三つ揃って、はじめて「使えるデータ」



何がどこまで決まっているのか

どの生物・組織・状態に由来するのか

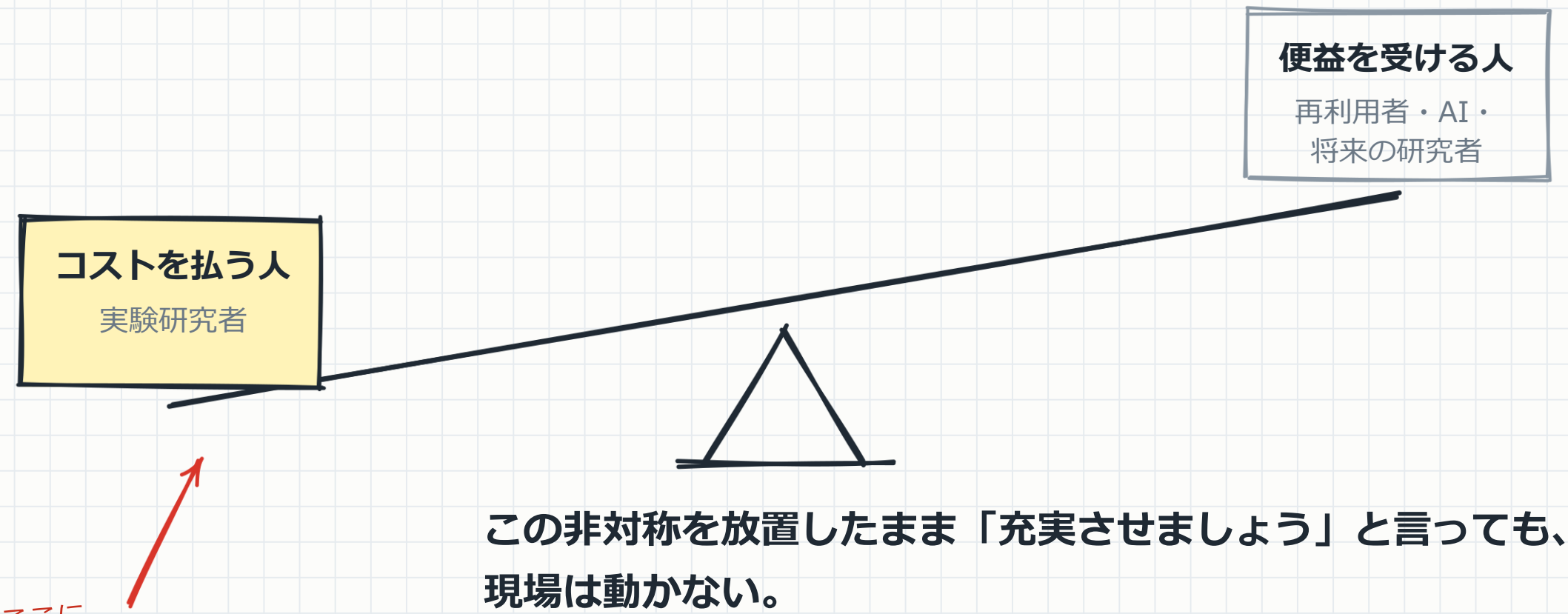
どう測り、どう処理した値なのか

「garbage in」の意味が変わった

間違ったデータではなく、
文脈を剥がされたデータのこと。

値は正しい。正しいまま、意味を失っている。

ところが、それを書くのは実験研究者



この非対称を放置したまま「充実させましょう」と言っても、
現場は動かない。

基盤構築の流れ

投稿の入口を一つに

GlycoPOST — 生データと実験条件を、同じ場所に同じ形で受け取る

「最低限」を先に決める

MIRAGE — 何を書けば再利用できるかを、分野として合意しておく

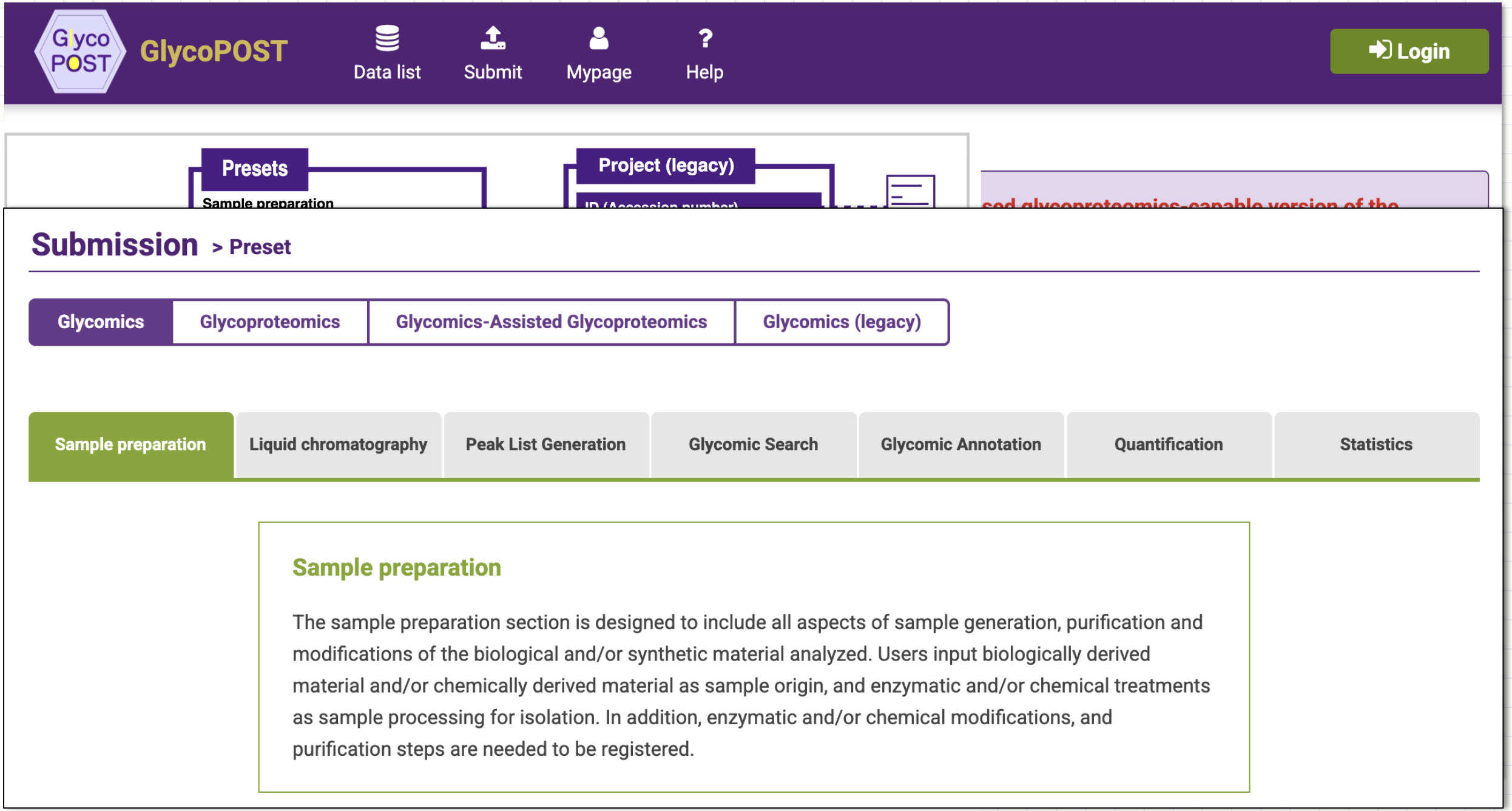
入力させず、拾い上げる

装置出力・解析ログから取れるものは、投稿者に打たせない

識別子で繋ぐ

GlyTouCan・GlyCosmos — 構造に ID を与え、他のデータと繋ぐ

「書いてください」から
「書かなくて済む」へ。


GlycoPOST
Data list
Submit
Mypage
Help
Login
Presets
Sample preparation
Project (legacy)
Submission > Preset
Glycomics
Glycoproteomics
Glycomics-Assisted Glycoproteomics
Glycomics (legacy)
Sample preparation
Liquid chromatography
Peak List Generation
Glycomic Search
Glycomic Annotation
Quantification
Statistics
Sample preparation
The sample preparation section is designed to include all aspects of sample generation, purification and modifications of the biological and/or synthetic material analyzed. Users input biologically derived material and/or chemically derived material as sample origin, and enzymatic and/or chemical treatments as sample processing for isolation. In addition, enzymatic and/or chemical modifications, and purification steps are needed to be registered.

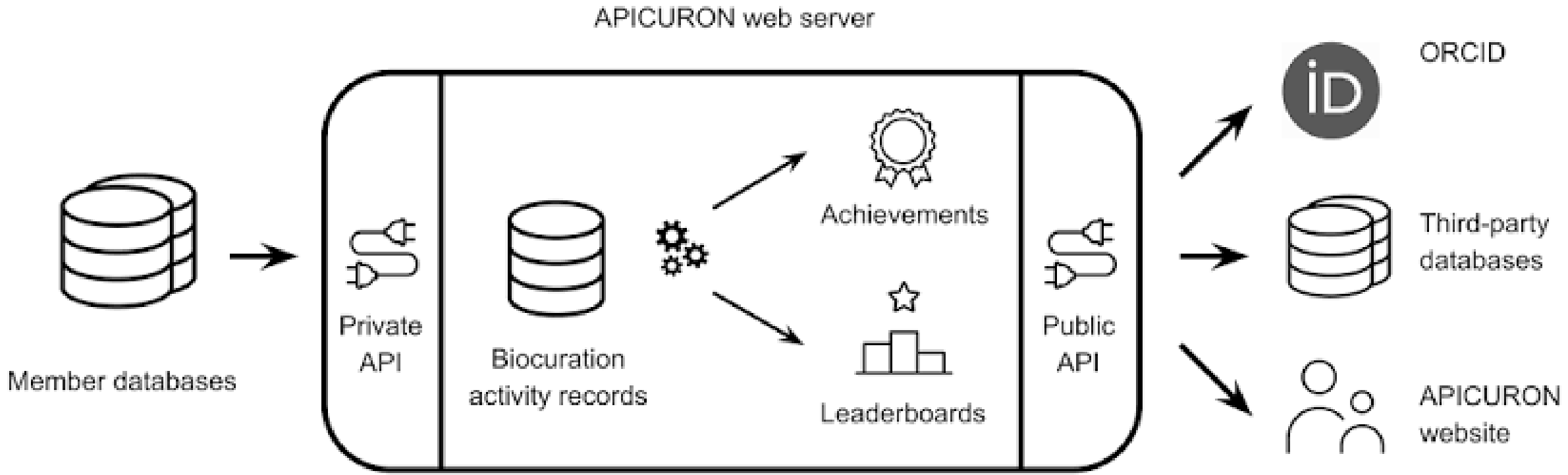
REPOSITORIES AND DATABASES

Data	MIRAGE Guidelines	Repository	Description
Glycans	**Sample Preparation**	**GlyTouCan**	Glycan structures and compositions
Glycomics MS Data	**Mass Spectrometric Analysis** **Sample Preparation**	**GlycoPOST** **UniCarb-DR**	MS, MS/MS, LC-MS, LC-MS/MS
Glycoconjugates MS data	**Mass Spectrometric Analysis** **Sample Preparation**	**GlycoPOST** **ProteomeXchange**	Glycoproteomics data only
Glycan array data	**Glycan Microarray Analysis** **Sample Preparation**	**GlyGen Glycan Array Data Repository** **CFG Glycan Array Experiments**	published and unpublished data archived experimental data
Liquid chromatography data	**Liquid Chromatography Analysis** **Sample Preparation**		No known public repository
NMR data	**NMR Glycan Recognition** **NMR Glycan Structures** **Sample Preparation**	**Biological Magnetic Resonance Database** **NMRShiftDB**	Repositories accept NMR data
Lectin array data	**Lectin Microarray Analysis** **Sample Preparation**	**LM-GlycoRepo**	
Capillary Electrophoresis data	**CE Analysis** **Sample Preparation**		No known public repository

議論したいこと

01 **「不完全でも大量に」**
vs.「少なくても文脈付きで」

AI 時代のデータベースは、どちらに
投資すべきか



AI が研究する時代に
人間にしかできない仕事は、
実験の文脈を残すことではないか。



木下 聖子　創価大学