

GlyCosmos RDFのアップデートとデータの拡充

○小野多美子¹、塩田正明¹、山田一作²、山本泰智³、木下聖子¹

1. 創価大学糖鎖生命システム融合研究所 (GalSIC), 2. 公益財団法人野口研究所, 3. 情報・システム研究機構ライフサイエンス統合データベースセンター (DBCLS)

INTRODUCTION

- リポジトリとデータベースを兼ね備える糖鎖科学ポータルとして、糖鎖、糖タンパク質、糖鎖関連遺伝子、疾患など、関連するデータベースを統合し、情報を提供することを目的としている
- 全てのデータは RDF で連携している

- 日本糖質学会により公式認定されたポータルサイトとして糖鎖生物学者への利便性の向上を図っている
- データリソースの更新、UIの利便性を高めるための更新など、4ヶ月に1回ポータルサイトのアップデートを実施している

GlyCosmos RDFのアップデート

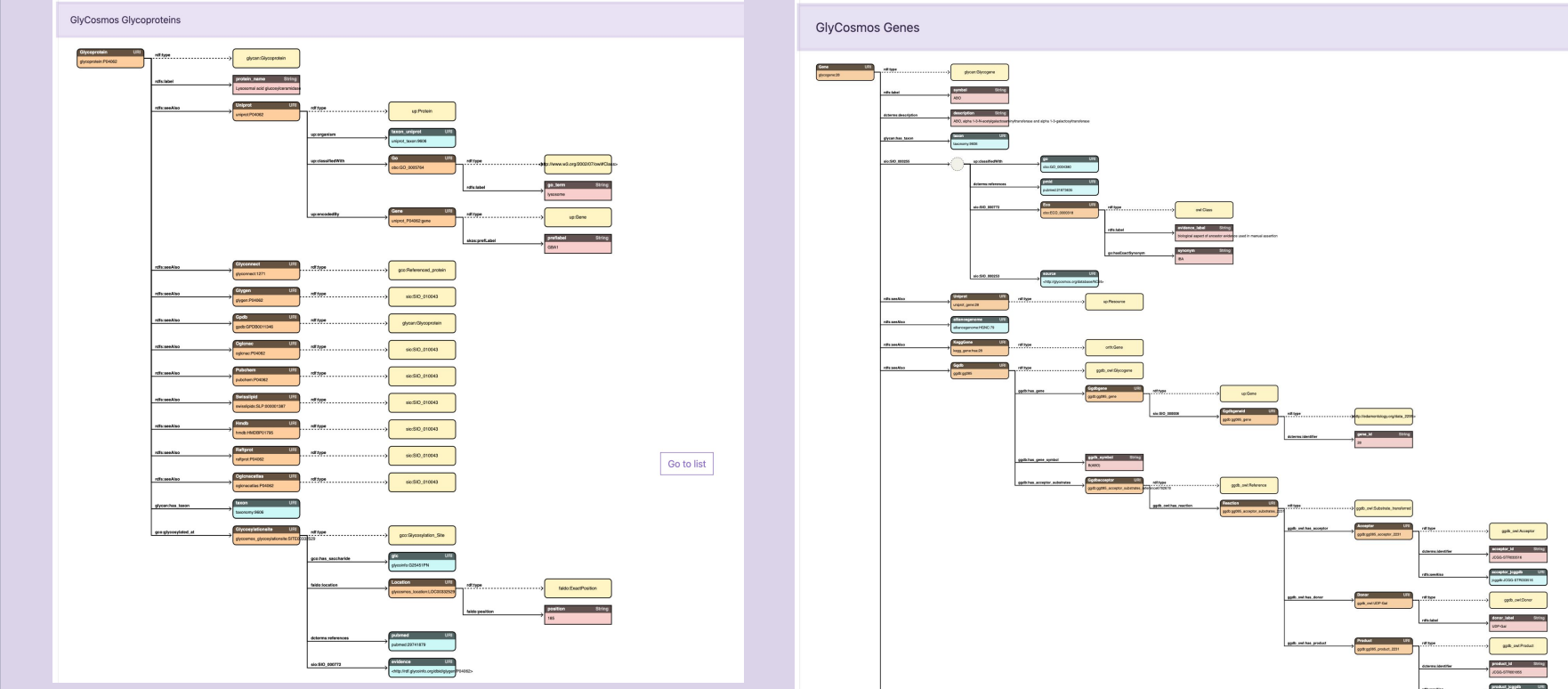
・GlyCosmosで作成したRDFスキーマ
 ・各DBが作成したRDFスキーマ
 それぞれのRDFスキーマの差異により

- 本来繋がるべきデータが繋がらない
- 煩雑なSPARQLクエリが必要

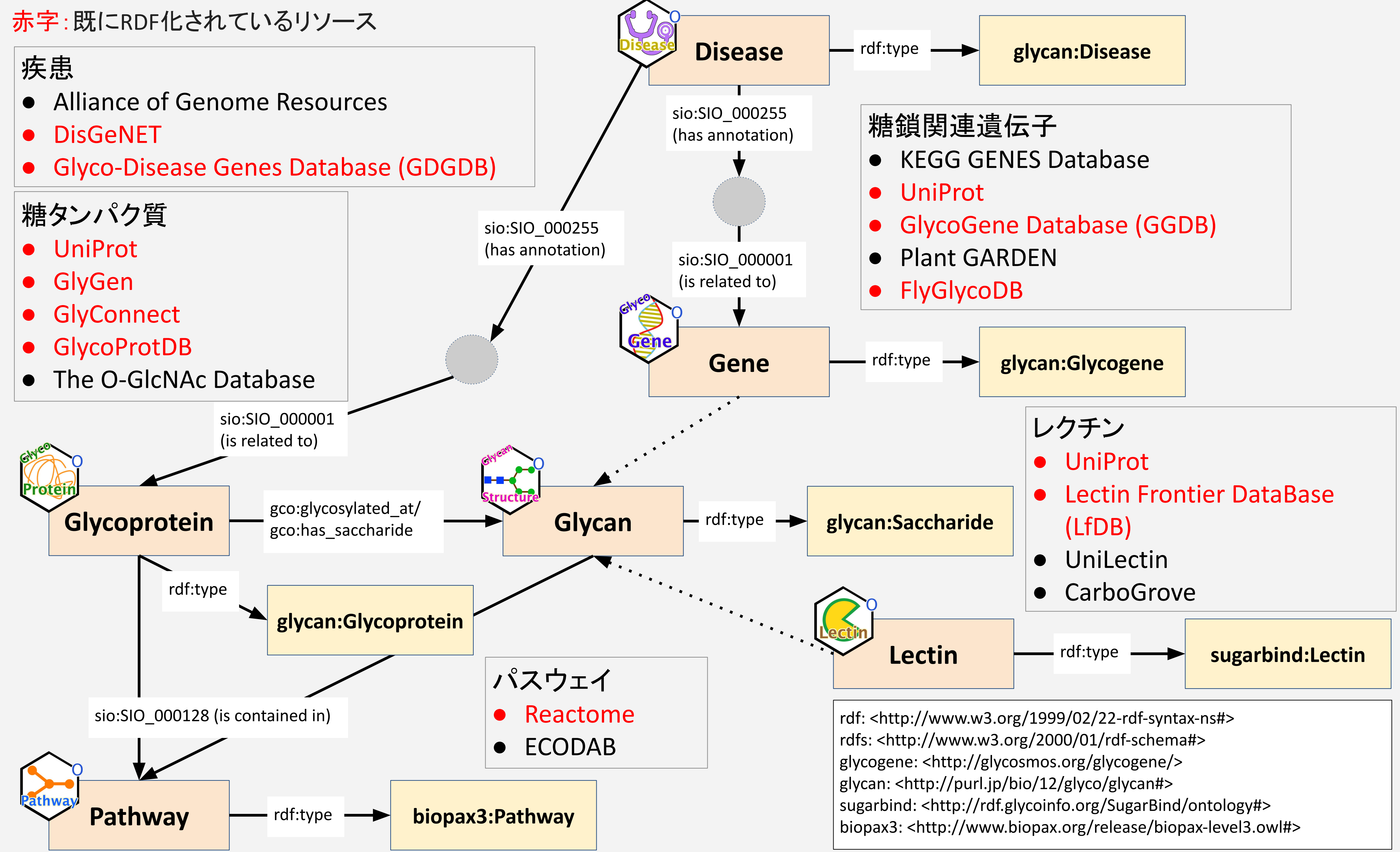
という問題があった。

GlycoRDF
Glycoconjugate Ontology (GlycoCoO)
Semanticscience Integrated Ontology (SIO)
 を活用し、GlyCosmos RDFとしてスキーマを作成した。

GlyCosmos RDFのスキーマ図は
GlyCosmos Programmatic Accessからアクセス可能。
 全てのスキーマ図はrdf-configで管理している。

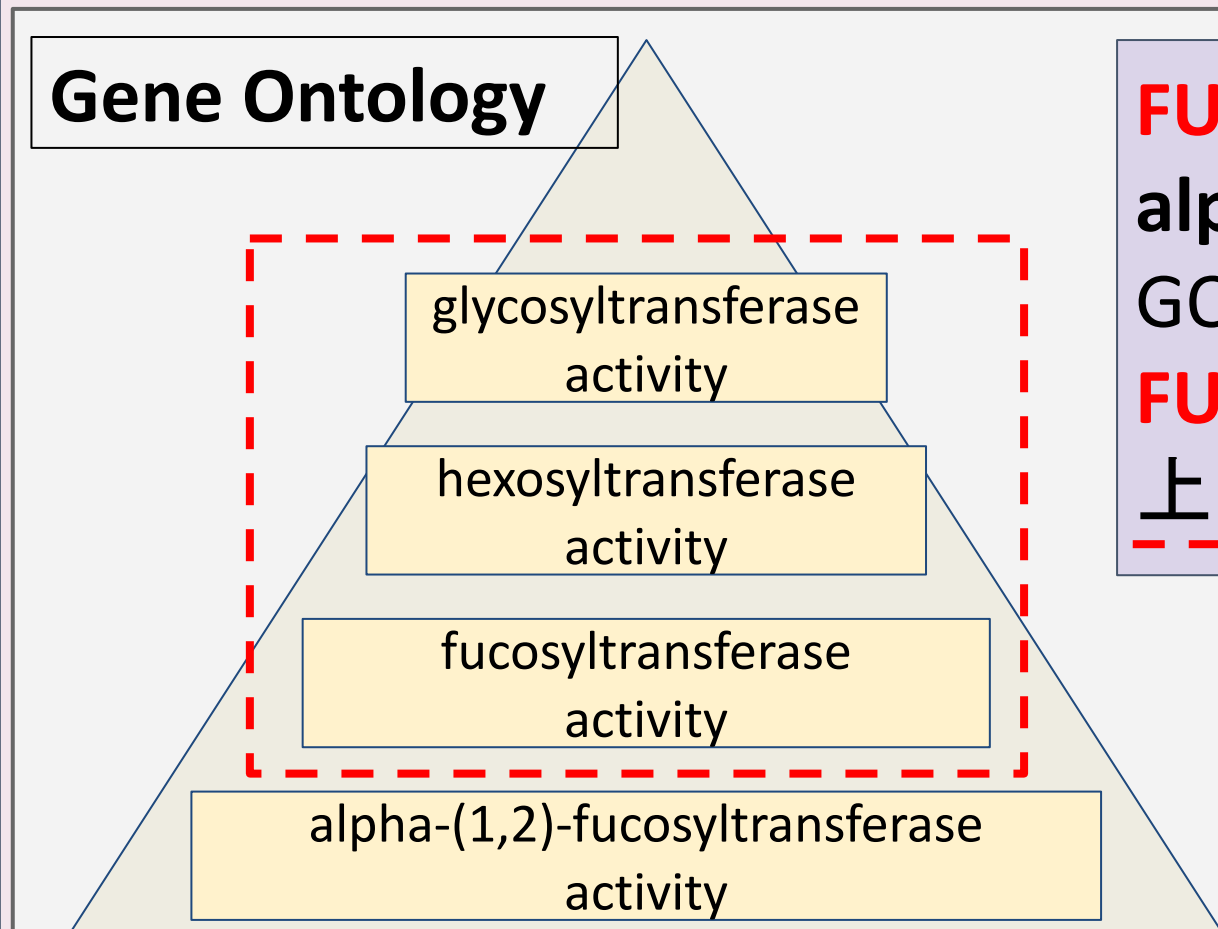


<https://glycosmos.org/programmatic>



Virtuosoの推論機能を用いたデータの拡充、検索機能向上

Virtuosoの推論について



FUT2遺伝子 (Gene ID:2524)に
alpha-(1,2)-fucosyltransferase activityという
 GO Termがついている場合、
FUT2には**alpha-(1,2)-fucosyltransferase activity**の
 上位階層にある機能も兼ね備える

推論を使うことにより、
 簡単なSPARQLクエリで上位階層のデータをま
 とめて取得できるようになる

GlyCosmos Pathway

RDFの改良と推論の活用
 により、糖タンパク質や糖
 鎖が関連するパスウェイ
 の検索が容易になった。
 (SPARQLの簡易化)

Beta-mannosidaseで検索

これまでは、階層の情報が取得できないため、**検索結果の表示が整理できておらず、ユーザーにとって見えにくい**表示であった。
 今回のアップデートで推論を活用することにより、**該当のタンパク質を含むパスウェイの上位階層のパスウェイが取得でき、ツリー構造での表示が可能**になった。

GlyCosmos Gene

Gene Ontology (糖鎖に関連する機能)で
 糖鎖関連遺伝子をフィルタリングできる機能を追加

glycosyltransferase activityで検索
 ↓
glycosyltransferase activityだけで
 なく、さらに下の階層のタームに該
 当する遺伝子も同時に検索が可能
 になった

上位階層の**glycosyltransferase activity**
 もSPARQLで取得できる

推論で取得

DISCUSSION

- GlyCosmosデータリソースの更新**
 - 糖鎖と糖鎖関連遺伝子・糖鎖とレクチンのデータについて、今後データが集まり次第スキーマ図を検討していく予定
 - そのために、様々な研究者の方々と協力してデータ収集を行っていきたい
- Virtuosoの推論機能の活用**
 - 糖鎖の部分構造検索の検索速度の向上
 - 糖鎖の質量、compositionと結合情報がわかる糖鎖との関係性 (subsumption) の検索

ACKNOWLEDGEMENTS

本研究は、科学技術振興機構 (JST)、バイオサイエンスデータベースセンター (NBDC) ライフサイエンスデータベース統合推進事業「統合化推進プログラム」の助成を受けています。

REFERENCES

- K. Aoki-Kinoshita, et al., J Biomed Semantics., 2013, 4(1):39. doi: 10.1186/2041-1480-4-39
- Yamada I, et al., Glycobiology, 2021, 31(7):741-750. doi: 10.1093/glycob/cwab065
- Dumontier M, et al., J Biomed Semantics. 2014, 5(1):14. doi: 10.1186/2041-1480-5-14
- Ashburner M, et al., Nat Genet., 2000, 25(1):25-9. doi: 10.1038/75556
- Gene Ontology Consortium, et al., Genetics., 2023, 224(1):iyad031. doi: 10.1093/genetics/iyad031
- Milacic M, et al., Nucleic Acids Res. 2024, 52(D1):D672-D678. doi: 10.1093/nar/gkad1025