# DBCLSにおける知識グラフによるバイオインフォマティクス研究



TogoDB publishes users' tabular data as a database and pr functionality to deliver the contents in RDF/SPARQL.

ine RUF Portal provides the world's largest collection of RDFs in life iciences and biomedical databases and serves SPARQL endpoints by utilizing SPARQL-proxy and RDF-config internally.

SPARQList creates REST APIs instantly by combining SPARQL queries ar JavaScript snippets for data retrieval and transformation in a simple

ID変換

eturns cached results for identical queries to improv . It also protects the SPARQL endpoint from DoS-like a query queuing and job management mechanism, a

五斗進、川島秀一、千葉啓和、細田正恵、池田秀也、三橋信孝、守屋勇樹、山本泰智、片山俊明

https://dbcls.rois.ac.jp/services

(DBLCS: 情報・システム研究機構 データサイエンス共同利用基盤施設 ライフサイエンス統合データベースセンター)

ライフサイエンス統合データベースセンター(DBCLS)では、生命科学系のデータベースの統合的利活用を促進するためのデータ基盤の構築と関連技術の 開発とともに、それらを用いたアプリケーションなどの開発を進めている。データ基盤としては知識グラフを用いた基盤構築を長年進めてきており、統合 化推進プログラムを含む各種データベースを開発者とも協力して知識グラフ化したデータセットをRDFポータルで公開している。現在、RDFポータルには オントロジーを含め各種生命科学の70以上のデータセットが収載されている。

RDFポータルには多様なデータが収載されており、それらを連携して利活用できるようにRDF化のガイドラインを整備するとともに、収載されたデータを 活用するためのミドルウェアも開発している。さらに、それらを用いたアプリケーションを提供し、データセットを組みわせて取得した結果をダウンロー ドしたり、ヒトゲノム変異情報を疾患などの情報と合わせて取得する仕組みも提供している。本発表ではDBCLSで開発している知識グラフ基盤とその応用

Data sources

R D F

√ JSON-LD

RDF

SPAROL

Proxy

GraphQL API

Analysis

Workflow

Grasp

CO

http://

41:42:41

SPARQL-proxy

SPARQL-GA

SPAROList

統合検索・マッピン

統合DBのショーケ

国際開発者会議BioHackathonやBiomedical Linked Annotation Hackathon (BLAH) の開催を行い、LLMによる知識グラフの利用技術やアノテーション技

国内版バイオハッカソン(一週間の泊まり込み)や Togothon(月一回2日連

続開催) などで国内の研究者・技術者の開発・情報交換の場を提供しています。

∞∞ 日本人バリアント

Knowledge graph infrastructure overview by T. Katayama

• RDF-config

Web Components

TOGO

Applications

国内・国際協力によるDB開発と標準化、

術など先進的な基盤技術開発を国際連携のもとで進めています。

国内データベース・リソース連携

RFST

について最近の話題とともに紹介する。

## 知識グラフ基盤構築によるDB統合とサービス開発

## 知識グラフの集約と提供

DBCLSがこれまでに培ってきた知識グラフによる統一的なデータ統合の技 術を発展させ、多くの研究者に利用されるサービスを提供することを目的 として、**データソース**の整備、データ整備・活用のための**ミドルウェア**の 開発、それらを基盤としたアプリケーションの開発を行っています。

## ータソース



着似子	DDBJ, HGNC, NOBI gene, NARO genebank	遺伝子オーソログ	HomoloGene, MBGD, OMA, PGDBj Ortholog
ゲノム	Ensembl, GTDB, GWAS catalog	タンパク質相互作用	InsAct, Instruct, HINT
メタゲノム	MicrobeDB.jp	パスウェイ・灰岩	BRENDA, REACTOME, WikiPathway, Rhea
エピゲノム	KERO, CNP-Mas, METHYL	システム生物学	BioModels, SSBD
ゲノム党員	Linked ICGC, Clinifer, ExAC, doSNP, doVer, dbNSFP, dbscNSV	原化	PRConto, GGDonto, DieGeNet, Clinitier, MedGen, CN/IC, COSMIC, TCGA, PubCaseFinder, KG-Hub/Covid
タンパク質	UniProt, neXtProt	パイオアッセイ	ChEMBL, PubChem
タンパク質2件構造	wwPCB, BMRB, FAMSBASE	用語集	MeSH, Allie, LSD, HGNC
min	GlyCoemos, GlyTouCen, GlycoEpilope, WURCS	トランスクリプトーム	ExpressionAfas, RefEx, KERO, Open TG-GATEs, Bgse
プロテオーム	(POSTdb, The Human Protein Atlas	メタボローム	MassBank, metabolonole
オントロジー	PMConto, GGDonto, NANDO, BioProtal, MeSH, OLS, HPO, GMOth 5-th	化合物	ChEBI, PubChem, Nikkaji, LingadBox, KNApSAcK, Polyinto, Nucleic Acid Drug
<b>メ</b> ク情報	TogolD, Quanto, Integbio DB catalog, NLM catalog, Coll, 新観線文レビュー	サンプル・細胞	BioSamples, JCM, NBRC, MediaDire, CelloSaurus, Growth Media
文献	PubMed, PubTator Central	(0.5.物・物物	BacDive, AgroLD, Plant GARDEN



アプリケーション

DiseaseSearch

エンドポイント情報収集・評価

## ミドルウェア

SPARQLクエリ作成



症状 - 疾患 - 遺伝子

TOGO MEDIUM 微生物培養培地

プロキシー **REST API** GraphQL API ウェブコンポーネント

# 非構造データとDBを統合利用するための技術開発

文献情報や画像情報など膨大な非構造データを、データベースの情報とと もに利用するための技術提供を行うことを目的として、AIやLLM技術も活 用しながら、文献情報アノテーションなどの基盤整備や技術・ツール開発 を行っています。



AT/IIM

•自然言語からの知識抽出

•自然言語問い合わせ

•正確な情報・翻訳語の提供



文献アノテーションレポジトリ テキストマイニング用辞書の レポジトリ

ライフサイエンス辞書

inMexes Allie Coll

**♦♦♦** NanbyoData

頻出表現・略語・引用関係

# ゲノムグラフを基盤とした解析技術の開発

最新のゲノムグラフ技術を基盤としたリファレンスゲノムの構築および解析サービス ヒトやNBRPのリソース生物種から環境微生物まで幅広く提供することを目指し ています。

- 公開可能な日本人リファレンスゲノムグラフ構築
- •スパコンでのゲノムグラフ解析環境の整備
- ・バリアント情報標準化研究会 (VISC)



# DB構築・キュレーションのための技術開発

日本のあらゆる生命医科学データを受け入れ、データマネージメントプランに基づい た適切なデータベース化を行うレポジトリ事業を推進することを目指し、データ登録 における効率化・自動化などレポジトリ間の連携に必要な技術開発を行っています。

•データ登録における効率化・自動化

DBCLSメンバーのみなさん

- •NBDCヒトデータベース
- •ヒト以外のデータ



ヒトデータ共有の Human data ためのレポジトリ



SNPアレイデータの インピュテーション解析

# **JPOS**

400

Togothon

**DICP** 









教育コンテンツ



NBRP

教育コンテンツ

-タベースやバイオインフォマティク スツールの動画マ ニュアル、講習会動画、



BodyParts3D/Anatomograph 解剖学的人体モデル

TOGOOTV

イラストの提供



JST/NBDC委託研究 LSI/NLDP(ライフサイエンス研究基盤事業ナショナルライフサイエンス . -タベースプロジェクト)

**NBDC** 

🧼 文部科学省

謝辞

**DBCLS** 

8 DDBI