

1 NBDCの基盤サービス紹介

トーゴーの日シンポジウム2021

○八塚 茂¹、信定 知江¹、宮崎 敦子¹、原菌 陸正¹、鈴木 拓真¹、片山 俊明²、川島 秀一²、畠中 秀樹^{1,2}、栗原 英輔³、加藤 健弘³、杉崎 太一郎⁴

1. 国立研究開発法人科学技術振興機構バイオサイエンスデータベースセンター(NBDC)、2. 大学共同利用法人情報・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)、3. 株式会社日立製作所、4. 三井情報株式会社

■ Integbioデータベースカタログ

Integbioデータベースカタログでは文部科学省、厚生労働省、農林水産省、経済産業省の4省による、生命科学系データベースの統合の一環として国内外の生命科学系データベースの所在情報と、データベースについての説明や生物種などのさまざまな属性情報(メタデータ)をまとめたリストを提供している。2021年8月末時点で2500件以上のデータベース情報を収録している。また、国内データベースの所在情報を広く国外に発信すると共に国外のデータベース情報を収集するため、国際的な生命科学系データベース情報を収集・公開しているFAIRsharingと2017年から連携し、データベース情報の交換を行っている。

The screenshot shows the Integbio database catalog interface. At the top, it says 'Integbio データベースカタログ' and '生命科学系データベースを一覧から探す'. There are search filters on the left, including '一覧内を検索する' and '一覧を絞り込む' with categories like '生物種' (Animalia, Plantae, Fungi, etc.) and 'タグ対象' (Genomes, cDNA, etc.). The main content area displays a list of databases with columns for 'データベースのレコード一覧 (件数: 2501)' and '並べ替え'. The list includes entries like 'PADS Arsenal: Prokaryotic Antiviral Defense System Arsenal', 'oRNAmot: oRNA motifs enrichment in complete transcriptomes', 'OGRDB: Open Germline Receptor Database', 'ncRNA-eQTL', 'ChemProteoBase: Proteomic profiling for classification of compounds', 'KNapsACK Biomarker', and 'Phos-tag SDS-PAGE データ集'. Each entry has a brief description and a '一括ダウンロード可能' button. On the right side, there is a 'メニュー' section with links like 'ホーム', '本カタログについて', and '更新履歴', and a '新着情報' section with recent updates.



1 NBDCの基盤サービス紹介

■ 生命科学データベース横断検索

生命科学データベース横断検索は、生命科学分野のコンテンツを対象にした検索システムで、分子DBの中を文献や特許情報とあわせて一括検索することができる。2021年8月末時点で、720以上のデータベースを検索対象にしている。

Cross search

What's New

- 2020.08.12 NLDB: Natural Ligand DataBase (カテゴリ: [タンパク質]-[ドメイン・モチーフ・立体構造]) を追加しました。
- 2021.08.12 「総合医薬」情報発信サイト (eJIM) (カテゴリ: [医薬・薬]-[医学]) を追加しました。
- 2021.08.03 人工知能研究センター デジタルヒューマン研究チーム 公開データ (カテゴリ: [文献]) を追加しました。
- 2021.07.21 龍典プラットフォーム (カテゴリ: [医薬・薬]-[医学]) を追加しました。
- 2021.07.05 TOHO rSNP DB (カテゴリ: [ゲノム・遺伝子・RNA]-[多型]) を追加しました。
- 2021.05.24 TOHO rSNP DB (カテゴリ: [ゲノム・遺伝子・RNA]-[多型]) を追加しました。
- 2021.03.31 KEGG NETWORK Database (カテゴリ: [メタボリズム]) を追加しました。
- 2021.03.31 核酸医薬品データベース (カテゴリ: [医薬・薬]-[薬学]) を追加しました。

統計情報

検索対象データベース数: 724DB
カテゴリ数: 42カテゴリ

■ 生命科学系データベースアーカイブ

生命科学系データベースアーカイブは、国内のライフサイエンス研究者が生み出したデータセットを長期間安定に維持保管し、データ説明（メタデータ）を統一して検索を容易にすると共に、利用許諾条件などの明示を行うことで、多くの人が容易にデータへアクセスしダウンロードを行えるようにしている。当アーカイブはNature publishingのScientific Data誌の推奨レポジトリに認定されている。2021年8月末時点で、152件のデータベースを収録している。

生命科学系データベースアーカイブ

アーカイブデータベース一覧 (ヘルプ)

データベース	データベース運用場所	代表者	データベースカテゴリ	生物種	要約 (キーワードを太字表示)	利用許諾
メタゲノム解析用リファレンス16S RNA配列データセット	ダウンロード	広島大学 大学院総合生命科学研究科 ゲノム情報科学研究所	核酸塩基	RNA配列データベース	Kraken2によるメタゲノムデータ解析で得られる16S RNA配列データセット	CC 表示 詳細
jPOSTデータベース	ダウンロード 詳細情報 オリジナルサイト	jPOSTチーム	石炭素	プロテオミクス	jPOSTデータベースは、jPOSTリポジトリに収録されたプロテオームデータに対して研究者が標準化した結果を収録したデータベースです。	CC0 詳細
LigandBox	ダウンロード 詳細情報 オリジナルサイト	近代化学技術科学研究所	塩素 硫黄 酸素	構造データベース/分子	計算機によるバーチャルスクリーニング用の分子化合物のデータベース	CC 表示-非営利 詳細



1 NBDCの基盤サービス紹介

トーゴーの日シンポジウム2021

■ NBDC RDFポータル

NBDC RDFポータルは、RDF形式の生命科学データベースを集積したポータルサイトである。本サイトでは、NBDCに寄託され「DBCLS RDF化ガイドライン」に基づいてレビューされたデータと、生命科学分野で重要なデータベースのミラーデータをダウンロード可能な形で公開する（リポジトリ）と共に、RDFのクエリ言語であるSPARQLで各データを検索することのできるサイト（エンドポイント）も公開している。

2021年8月末時点では、53のRDFデータセット、約1010億トリプルを公開している。

公開日	データセット名	説明
2021-08-04	PubChem	PubChemは、NIHが提供する、網羅的な化合物のデータベースです。詳細な情報は次のURLを参照してください: https://pubchem.ncbi.nlm.nih.gov/ ● 薬物/化学物質 ● その他の生体分子
2021-06-30	IDC	Integbioデータベースカタログ/RDF Integbioデータベースカタログ/RDFはIntegbioデータベースカタログデータをRDFへ変換したものです。 ● その他 ● その他
2021-05-18	PubCaseFinder RDF	PubCaseFinderデータベースのうちの、ウェブリソースおよび症例報告から抽出した疾患-症状関連データおよび遺伝子 ● 健康/疾患 ● 遺伝子 ● その他
2021-04-07	UniProt	The Universal Protein Resource (UniProt) は、タンパク質配列と機能注釈に関する網羅的なデータベースです。詳細な情報は https://www.uniprot.org/ を参照してください。 ● 蛋白質 ● 配列
2021-01-20	BMRB/RDF	BMRB/RDFはNMR-STARデータをRDFへ変換したものです。 ● 蛋白質 ● その他の生体分子 ● その他 ● 配列 ● 構造
2021-01-20	wwPDB/RDF	wwPDB/RDFはPDBx/PDBMLデータをRDFへ変換したものです。 ● 蛋白質 ● 薬物/化学物質 ● その他の生体分子 ● 配列 ● 構造
2020-11-11	NANDO	難病オントロジーNANDOは、難病に関する語彙や、難病と他の疾患データリソースとの関連を提供します。NANDOに基づいて構築されたオントロジー/用語/学名/命名法 ● 健康/疾患 ● タグ配列 (核酸) ● オントロジー/用語/学名/命名法
2020-08-31	NBRC	NBRC微生物株データセットとは、NBRC (NITE Biological Resource Center) で維持・管理・分譲されている微生物株のデータ ● 個体/種 ● 細胞 ● バイオリソース
2020-06-10	DBKERO RDF	DBKEROは細胞のマルチオミクスデータ (SNV, RNA-seq, ChIP-seq, BS-seq, TSS-seq) を集めたデータセットです。ChIP-seq ● ゲノム ● 多型 ● その他のDNA ● 遺伝子発現 ● その他
2020-06-09	jPOST database RDF	jPOST database RDFはjPOSTプロジェクトにおいて再解析されたプロテオームデータセットについて記述したものです ● 蛋白質 ● 配列