

別紙1 構築、公開したデータベース等

研究代表者名: 小原雄治
 研究開発課題名: データベース統合に関わる基盤技術開発

名称	概要	URL	公開日	レコード数	レコード数の定義、説明等	NBDCアーカイブへのデータ提供	備考
PhylogeneticTree (DNA, RNA, Protein) Active ワークフロー	解析結果を基に自動でレポート作成する機能を追加しました。	http://togo.cbrc.jp/active_local_phylo.html	2013年12月16日				最終アップデートの日: 2013/12/16
Protein Structure Prediction Active ワークフロー	タンパク質関連解析ワークフローにデンマーク工科大学(DTU)の解析ツールを加えました。	http://togo.cbrc.jp/active_local_rna_prediction.html	2013年4月22日				最終アップデートの日: 2013/9/17 外部解析ツールとしてDTUの14ノードの追加
RDF化したライフサイエンス辞書のSPARQLエンドポイント	ライフサイエンス辞書プロジェクトにより編纂された辞書をRDF化し、SPARQLで問合せ可能とした。	http://purl.jp/bio/10/lis/sparql	2013年2月1日	6340945	トリプル数		H24年度公開
RNA Structure Prediction Active ワークフロー	RNA構造予測ワークフローに立体構造を考慮した相互作用解析ツールを加えました。	http://togo.cbrc.jp/active_local_rna_prediction.html	2013年1月17日				最終アップデートの日: 2013/9/17 独自2解析ツールを追加
Molecular Simulation Active ワークフロー	タンパク質のモデリング、ドッキング、分子動計算を全て行うワークフローです。外部連携ノードやドッキングサイトを指定できるノードを追加しました。	http://togo.cbrc.jp/active_local_mol_simulation.html	2012年9月28日				最終アップデートの日: 2013/12/27 外部DB連携ノードとして4ノード、外部の4解析ツール、独自4解析ツールを追加
TogoTable	データベースIDを含む表形式のデータに対して、IDを検索キーにして複数のトリプルストアからデータベースをまたいだアノテーション情報を取得するシステムです。	http://togotable.dbcls.jp/	2012年9月3日				平成23年度プロトタイプ開発。平成24年度ベータ版公開。平成25年度コンフィグレーション機能追加
高速配列検索 GGGenome	塩基配列を高速に検索するウェブサーバーです。	http://GGGenome.dbcls.jp/	2012年7月4日	174596856	塩基配列の数	対象外	RefSeqの更新に合わせて2ヶ月に1回更新
反応オントロジー	代謝系で使われている酵素反応から、基質生成物ペアを抽出し、反応パターンを定義したデータベースRCLASSと、RCLASSに基づいた反応分類をまとめて公開しています。	http://www.genome.jp/reaction/	2012年7月1日	17,366	基質化合物ペアと反応パターンの合計数 (2014.2.28現在)	未対応	ゲノムネット ケミカル情報解析ツールの項参照
Semantic service(概要)	セマンティック技術に対応したSADI開発を紹介しているサイトです。	http://togo.cbrc.jp/semantic.html	2012年4月20日				最終アップデートの日: 2014/2/13 日本語・英語版HELPページの修正
Semantic service(解析ツール)	これまで開発した解析ツールをSADI仕様に従い公開している解析サービスサイトです。	http://semantic.cbrc.jp/sadi-services/	2012年4月20日	28	14個のSADI解析サービスを2つのオントロジーを利用し公開(CBRCオントロジーとSIOオントロジー)		最終アップデートの日: 2014/9/23

名称	概要	URL	公開日	レコード数	レコード数の定義、説明等	NBDCアーカイブへのデータ提供	備考
ライフサイエンス 領域融合レビュー	生命科学において注目される分野・学問領域における最新の研究成果について、第一線の研究者の執筆による日本語のレビューを、だれでも自由に閲覧・利用できるよう、公開するオンラインジャーナルサービスです。平成25年1月末までの17か月間に24本のレビューを公開しました。	http://leading.lifesciencedb.jp/	2012年1月9日	24	レビューの数	対象外	2013年度にウェブサイト機能を改良(関連記事表示、閲覧数ランキング表示、インクリメンタル検索)
統合遺伝子検索 GGRNA	遺伝子や転写産物をさまざまなキーワードからすばやく検索し、その結果をわかりやすく提示することができる遺伝子検索エンジンです。遺伝子名や各種ID、タンパク質の機能や特徴などのキーワードだけでなく、短い塩基配列やアミノ酸配列から遺伝子を高速に検索することも可能です。	http://GGRNA.dbcls.jp/	2011年5月18日	5760653	塩基配列の数	対象外	RefSeqの更新に合わせて2ヶ月に1回更新
DBCLS SRA	公共データベース(SRA [NCBI]、ENA [EBI])、DRA[DDBJ])に登録された「次世代シーケンサ」データについて、目的別、機器別、生物種別等、さまざまな統計情報から閲覧、比較、データのダウンロードができる目次サイトです。論文からのデータの検索も可能です。	http://sra.dbcls.jp/	2011年5月1日	543465	プロジェクト、実験、文献数		週1回データ更新。1年でデータ数は倍々ペース。2013年度にKusarinokoとして提供していたメタデータ検索システムを統合し全文検索エンジンを搭載。
Allie RDF Data	Allie のSPARQL エンドポイントです。	http://data.allie.dbcls.jp/	2011年1月12日	120130005	トリプル数		本プロジェクトでの新規開発。コンテンツは毎月更新。
Semantic TogoDB	TogoDB の機能に加え、アップロードしたデータを半自動的にRDFへ変換する機能をもつシステムです。	http://togodb.org/	2011年1月12日	18430715	ユーザの作成した687データベースの総計		初期版のTogoDBにセマンティック・ウェブ対応を行ったプロトタイプ
RefEx (Reference Expression dataset)	EST、GeneChip、CAGE、RNA-seq の4種類の異なる手法によって得られたヒトおよびマウス、ラットにおける遺伝子発現データを並列に表示し、遺伝子発現解析を行う上で基準となるリファレンス(参照)データベースとして利用することを目的とした遺伝子発現データベースです。	http://refex.dbcls.jp/	2011年1月10日	213993	ID重複のない発現データ数		2012年度にインターフェース刷新。GeneChipデータ更新。2013年度にRNA-seqデータ更新。検索機能拡充。
togo picture gallery	ライフサイエンス分野のイラストをだれでも自由に閲覧・利用できるようWeb上にて無料で公開しているウェブサイトです。	http://g86.dbcls.jp/togopic/	2011年1月4日	129	イラストの数		2013年は41増

名称	概要	URL	公開日	レコード数	レコード数の定義、説明等	NBDCアーカイブへのデータ提供	備考
BioHackathon	最先端の研究開発者を招聘した国際的なソフトウェア開発会議の情報交換ならびに成果を公開するためのサイトです。BioHackathon はH19 年度からウェブサービス標準化、統合利用環境構築、セマンティックウェブ、リンクトデータで4 度開催しています。DB 統合利用に関する内外研究者の国際的ハブとしての地位も確立しつつあります。	http://www.biohackathon.org/	2011年1月1日				BioHackathon開催時に利用
LifescienceQA	ライフサイエンス分野の研究者の情報交換を促すためのQ&A サービスです。現在までに116 の質問が投稿され、224 の回答が得られています。	http://qa.lifesciencedb.jp/	2010年1月11日				システム更新なし
ライフサイエンス 新着論文レビュー	Nature、Science、Cell などのトップジャーナルに掲載された日本人を著者とする生命科学分野の論文について、論文の著者自身の執筆による日本語によるレビューを、だれでも自由に閲覧・利用できるよう、いち早く公開するオンラインジャーナルサービスです。平成25年1月末までの41 か月間に601 本のレビューを公開しました。	http://first.lifesciencedb.jp/	2010年1月9日	601	レビューの数	対象外	2013年度にウェブサイト機能を改良(関連記事表示、閲覧数ランキング表示、インクリメンタル検索)
TogoDoc	文献情報及び論文PDF を管理し、また、特定の文献情報群に関連する論文情報を提示するシステムです。TogoDoc Client と連携して文献を管理することも可能なほか、スマートフォンにも対応しています。	https://docman.dbcls.jp/pubmed_recom/	2009年1月12日	284102	利用者ごとの登録文献や関連語数		iOS向けクライアント開発。平日毎日PubMed更新に合わせた推薦用のデータ更新。
DBCLS galaxy	生命科学データに特化したウェブベースの対話的ツール組み合わせインタフェースです。DBCLSで開発されたツール群も組み込んでいます。	http://galaxy.dbcls.jp/	2009年1月10日				H25年度セマンティック・ウェブのフレームワークSADIIに入出力を対応させた。
TogoWiki	国内版バイオハッカソンの情報交換ならびに成果を公開するためのサービスです。	http://wiki.lifesciencedb.jp/mw/	2009年1月8日				国内版Hackathon、Sparqlthon 開催時に成果まとめサイトとして利用
inMeXes	MEDLINE を対象とし、利用者が入力した文字列とマッチする表現を逐次的(1文字入力毎)に検索するシステムです。論文中の英語表現を容易に検索できます。	http://docman.dbcls.jp/im/	2009年1月7日	61858137	検索対象nグラム数		関連研究分野の追加、正規表現フィルタの追加
Gendoo (Gene, Disease Features Ontology-based Overview System)	文献情報をもとに、遺伝子、疾患について、関連する疾患、薬剤、臓器、生命現象などの特徴をキーワードでリスト表示するツールです。	http://gendoo.dbcls.jp/	2008年12月12日	11537465	遺伝子/疾患と関連キーワードの組数		原則、年一度、データ更新。対象生物種も3から9に拡充

名称	概要	URL	公開日	レコード数	レコード数の定義、説明等	NBDCアーカイブへのデータ提供	備考
Allie	MEDLINE を対象とし、出現する略字とその正規系のペアを検索するシステムです。略字を入力することで、その使われ方を一覧表示します。	http://allie.dbcls.jp/	2008年1月4日	101943338	略語や展開形、関連文献情報、関連略語情報などの総数		インターフェースの向上、日本語対訳の充実、表記上の揺れをまとめるクラスタリングなど。コンテンツは毎月更新。
DBCLS OpenID	一つのID で複数のサイトを認証できるシステムです。各サイトで認証サービスを用意する必要がなく、サイト間のユーザー情報の集約が容易におこなえます。	http://openid.dbcls.jp/	2008年1月4日				システムの更新なし。運用のみ。
TogoDB	エクセルなど表形式のデータを簡単に読み込み、DB 化し、自動的に共通のウェブ検索インタフェースを生成するシステムです。	http://togodb.dbcls.jp/	2008年1月4日	202032927	ユーザの作成した777データベースの総計		Semantic TogoDBの機能を取り込みオントロジー設定などの高機能化
TogoWS	国内外のウェブサービスを共通のAPIで利用できる仕組みと、サービス間の連携に必要なデータ形式変換機能、サービスの稼働状況の監視等を提供するシステムです。	http://togows.dbcls.jp/	2008年1月4日				RDFへの変換機能、UCSCのヒトゲノムに対応などの機能追加
MotDB	講習会の開催情報、参加受付ならびに講習会で使用する資料等を配布するためのサイトです。平成23年度は6回の講習会を開催しました。	http://motdb.dbcls.jp/	2008年1月2日	44	講習会44回分		講習会開催時にページ追加
統合TV	生命科学分野の有用なデータベースやウェブツールの活用法を動画で紹介するウェブサイトです。	http://togotv.dbcls.jp/	2007年7月19日	757	動画の数	対象外	2011年度は157本公開。2012年度は61本公開。2013年度は107本公開。

名称	概要	URL	公開日	レコード数	レコード数の定義、説明等	NBDCアーカイブへのデータ提供	備考
BodyParts3D	BodyParts3D(ボディパーツ3D)は人体各部位の位置や形状を3次元モデルで記述したデータベースです。3Dレンダラー上のウェブAPIを使うモデルエディター(アナトモグラフィー)を使って、BodyParts3D から解剖概念を選択して自由に人体のモデル図を作成、交換でき、利用者の情報もモデル上にマップ表現できます。データのバージョンアップに加えデータの管理方法を刷新。概念名をファイル名称とする旧方式からユニークなファイル名称と概念番号の対応表で持つようにした。これによって複数のバージョンを同時に提供可能になり、造形担当者とデータ(オブジェクトファイル)と概念の対応を行うデータ管理者の作業の分離で効率化。これにともないデータベースサイト(データレジストリ)も変更され概念とモデルデータを明確に区別し混乱回避。	http://lifesciencedb.jp/bp3d/	2007年5月10日	Elemental:1652 Compound:1854 (ver3.0 2011、 Juneでは E:915,C:1053)	単位は概念表現数。 Elemental 表現=3Dオブジェクトファイル数。単独のオブジェクトでの表現をelemental(representation)、これらを組み合わせて行う表現をcompound (representation)と呼んでいます。	提供済み	2012年6月にversion4.0を公開
OReFiL	オンライン上に存在する多数の生命科学系の資源(データベースやソフトウェアなど)を効率的に見つけるための検索システムです。	http://orefil.dbcls.jp/	2007年1月8日	97937	資源数及び関連論文数		コンテンツは毎週更新、機能追加は無し。
ゲノムネット ゲノム情報解析ツール	ゲノムから遺伝子の機能アノテーションを中心としたゲノム情報を解析するためのツール群です。遺伝子の自動アノテーションとパスウェイマッピングのためのKAAS、KEGG MODULE を使った機能割り当て解析ツール MAPLE、EST からコンセンサスコンティグを作成するEGassembler、オミックスデータ統合による遺伝子ネットワーク予測ツール GENIES から成り立っています。	http://www.genome.jp/ja/gn_tools_ja.html#genome	2005年7月7日			未対応	この3年間で新たに、自動計算で構築したオーソログクラスタOCを検索できるようにするとともに、メタゲノムやゲノムをアノテーションし KEGG MODULE を用いて解釈するためのシステム MAPLE を公開した。 また、KAASをメタゲノムに対応するために生物種の割り当てを行えるように拡張するとともに、GENIESのサンプルデータを増やして使い勝手を改良した。
ゲノムネット ケミカル情報解析ツール	代謝化合物や反応を中心としたケミカル情報を解析するためのツール群です。類似化学構造検索ツールSIMCOMP、部分構造検索ツールSUBCOMP、反応経路計算・予測ツールPathComp、PathPred、PathSearch、EC 番号割り当てツールE-zyme から成り立っています。	http://www.genome.jp/ja/gn_tools_ja.html#chemical	2004年10月1日			未対応	この3年間では、反応オントロジーで構築した反応分類情報を用いた、類似反応パス検索システム PathSearch を開発して、公開した。また、反応分類や新たに開発した化合物構造フィンガープリントを用いて E-zyme を改良するとともに、新規反応を遺伝子と結び付ける手法を開発し実装した。これは E-zyme2 として2014年4月に公開予定である。

名称	概要	URL	公開日	レコード数	レコード数の定義、説明等	NBDCアーカイブへのデータ提供	備考
LinkDB	世界各国で開発されている分子生物学関連のデータベースをエントリー間の関係で結んだデータベースです。データベースエントリー間の関係とその属性(original, reverse, equivalent)の3項関係で表現しています。KEGGを中心に162のデータベースを約25億のリンクで結んでいます。	http://www.genome.jp/linkdb/	1994年11月26日	2,555,806,094	エントリー間の関連(2項関係)の数(2014.2.28現在)	未対応	この3年間では、全データをRDF化するとともに、REST形式のAPIで検索できるようにした。また、Virusoを用いてSPARQL検索できるようにするとともに、Stanzaを用いた検索例を作成した。さらに、LinkDBのホームページを大幅に改訂し、データベースリンク図からデータベース間のリンク情報をダウンロードできるようにするとともに、各データベースのURIリストなどを参照できるようにした。データの更新は継続して行うとともに、新たに ChEMBL、SIDER、UM-BBD、SABIO-RK、PGDBj、RIKEN-BRC、OCへのリンク情報を追加した。
PubAnnotation, TextAE	Webによる文献アノテーションのための統合環境。PubMed, PMCのオープンアクセスな文献に関しては文献データのファイルフォーマットを標準化し文字列の絶対番地を提供することによって、異なるグループの文献アノテーションを一元的に取り扱うことができえる。遺伝子名辞書などにより自動で固有表現抽出が可能	http://pubannotation.org/					アルファサービス中
OntoFinder/OntoFactory	データをRDF化する際に適したオントロジーの検索と推薦、マッピングをするシステム	http://ontofinder.dbcls.jp/					アルファサービス中
LODQA	自然言語による質問応答システム。自然文で書いた質問からSPARQLを自動生成しRDFデータの検索が可能。	http://lodqa.org/					アルファサービス中
PubDictionaries	辞書データをウェブで共有し、辞書に基づくテキストアノテーションが行えるシステムです。	http://pubdictionaries.org/					アルファサービス中