National Bioscience Database Center

独立行政法人 科学技術振興機構 Japan Science and Technology Agency

生命科学分野のデータベースに関する情報集約に向けて ～カタログ構築の現状

○坂東明日佳[1]、河村明子[1]、大久保克彦[2]、熊谷禎洋[2]、川本祥子[3]
1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、2(株)日立ソリューションズ、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

NBDCのカタログサービスとは？

データベース統合化の段階的進展

平成25年度までに実施予定

(1)データベースのカタログ化
ポータルサイトに、各省のデータベースに関する情報を記載・登録し、各省のデータベースに対して、**リンクが行われる状態**を実現する。

(2)横断検索連携
各省における適切なサイトに、横断検索サーバーを設置するとともに検索インデックスを統一化し、各省のデータベースに対して、**相互に一括横断検索が可能となる状態**を実現する。

(3)アーカイブ構築連携
新センターのポータルサイトに、各省のデータベースをガイドラインに沿ってアーカイブ化して保存し、各省のデータベースに対して、**統一形式でのデータダウンロードが可能となる状態**を実現する。

平成25年度以降

(4)データベース構築連携
新センターのポータルサイトに、各省から譲り受けたデータベースを再編して収容し、各省のデータベースが、**統一化したフォーマット で再 構築 された状態** を実現する

NBDCが提供するデータベースカタログサービス

生命科学系データベースカタログ（DBカタログ）
http://biosciencedbc.jp/dbcatalog/
892件公開中（更新データ916件）*。収録データベース数を重視。

WINGPro
http://wingpro.biosciencedbc.jp/dbpwiki/
417件*公開中。データベースの使い方を含む項目立て。
SEO対策重視。

＊2011年9月26日現在

今年度の主な課題

●カタログの項目と内容の整備：カタログ間の情報を集約し、サービスに一体化を持たせる
●各省のデータベース統合事業との連携：各事業の持つカタログ情報の取り込み

カタログの項目と内容の整備

1. 関連レコードの突き合わせによる、相互リンクの実現

	公開件数	相互対応あり
DBカタログ	916 件	258 件
WINGPro	417 件	265 件

WINGpro: http://wingpro.biosciencedbc.jp/dbpwiki/index.php/Database for the Exchangeable Gene Trap Clones (EGTC)

生命科学系データベースカタログ http://biosciencedbc.jp/dbcatalog/dbcatalog.cgi?i=EGTC

DBカタログ　相互リンク　WINGPro

各カタログサイト上で相互に関連するレコード情報を容易に参照可能に。

2. NBDC統一データベースIDの作成と付与

IDの使用目的：NBDC内でのデータベース情報管理のために使用するID

IDフォーマット： Prefix【NBDC】＋数字5桁　【例】NBDC00001

NBDC ID	DB名_Short Name	DB名_Long Name	URL	DB名	URL
			DBカタログの項目	WINGproの項目	
NBDC00123	eF-site	electrostatic surface of Functional-site	http://ef-site.protein.osaka-u.ac.jp/eF-site	EF-site	http://ef-site.hgc.jp/eF-site/
NBDC00124	EGTC	Database for the Exchangeable Gene Trap Clones	http://egtc.jp/action/main/index	Database for the Exchangeable Gene Trap Clones (EGTC)	http://egtc.jp/view/index
NBDC00125	Caenorhabditis elegans WWW Server		http://elegans.swmed.edu	エントリーなし	エントリーなし
NBDC00126	The Multi-Dimensional Human Embryo		http://embryo.soad.umich.edu/index.html	エントリーなし	エントリーなし

・アーカイブ、横断検索等サービス間でのデータ交換や相互リンクの促進
・各省とのデータベース情報の共有をスムーズに

4. 各項目の記載内容の整備（準備中）

DBカタログ、WINGProのコンテンツを元に新しく定義した項目の記載仕様に沿って記述内容を整備していく。

	DB名	説明
DBカタログ	Database for the Exchangeable Gene Trap Clones	DBカタログ：熊本大学発生医学研究センター臓器形成分野で作製した可変型遺伝子トラップクローンのデータベースです。トラップした遺伝子に関する情報や5'-RACE法による同定情報、トラップクローンから樹立したマウス系統情報を掲載しています。(後略)
WINGPro	Database for the Exchangeable Gene Trap Clones (EGTC)	遺伝子置換の可変型遺伝子トラップクローンのデータベース

新DB名　新説明

3. メタデータ項目の整備

既存のデータベースカタログサービスの項目比較

WINGpro（エントリー）	WINGpro収録一覧	WINGpro 構築型分類完全版	DBカタログ	DBカタログ（更新版）	生命科学系データベースアーカイブ	MEDALS データベース使録	BioDBCore
1.タイトル	1.名称	1.名称	1.#DBID	1.LSDBID	1-1.名称	1.成果物名	1.Database name
2.別称	提供サイト *	提供サイト*	2.URL	2.ID	1-2.読み方	2.成果物の別名	2.Main resource URL
3.提供機関	2.別称	2.Webページのタイトル	3.タイトル(日本語)	3.言語	1-3.別名	3.成果物に関する説明	3.Contact information (e-mail; postal mail)
4.提供機関の所在(国)	3.概略	3.概要	4.タイトル(英語)	4.URL	1-4.作成者	4.成果物のタイプ	4.Date resource established (year)
5.URL	4.提供機関	4.開発機関	5.LsdbName(日本語)	5.DB名（short）	1-5.連絡先	5.運用機関	5.Conditions of use (Free, or type of license)
6.カテゴリ	5.開発機関	5.情報源	6.LsdbName(英語)	6.DB名(long)	1-6.データベース分類	6.機関所在国	6.Scope: data types captured, curation policy, standards used
7.提供データ	6.国名	6.国名	7.説明(日本語)	7.日本語タイトル	1-7.生物種	7.サイトURL	7.Standards: MIs, Data formats, terminologies
8.取り扱い生物種	7.型分類	7.型分類	8.説明(英語)	8.概要（日本語）	1-8.説明	8.インターフェイス	8.Taxonomic coverage
9.提供様式	8.対象	8.対象	9.分類	9.概要（英語）	1-9.データベースの特長•有用性•活用方法	9.入力例	9.Data accessibility/output options
10.利用条件	9.生物種	9.生物種	10.カタログ	10.上位データベース	1-10.予算的背景•プロジェクト	10.キーワード	10.Data release frequency
11.データ量	10.提供様式	10.PubMed ID	11.組織	11.下位データベース	1-11.論文等	11.ダウンロードデータ総量（Mbyte)/データ一括取得方法	11.Versioning period and access to historical files
12.更新頻度		11.カタログサイト掲載数	12.更新日時	12.DB型	1-12.オリジナルサイト	12.使っている外部リソース	12.Documentation available
13.更新記録		12.最終アクセス確認日	最終アクセス日時	13.対象	1-13.運用開始年月日	13.主な対象データ	13.User support options
14.説明（項目なし）		13.内容		14.和名	1-14.最終更新年月日	14.生物種	14.Data submission policy
15.利用方法		14.説明		15.一般名	1-15.統括サイト	15.利用条件	15.Relevant publications
16.Tips		15.参考URL/PubMedID		16.学名	1-16.データの一括ダウンロードサイト	16.データ更新頻度（過去2年間）	16.Resource's Wikipedia URL
17.その他		16.素材となったDB**		17.taxID	1-17.参照先データベース	17.最終更新日（調査日）	17.Tools available
		17.データ量		18.Reference(PMID)	1-18.一覧表示	18.利用できるID	
		18.PubMed ID（手動収集）		19.Reference（その他文献）	1-19.クエリ検索	19.IDを使った成果物の利用方法	
		19.統計情報掲載ページ		20.Reference（サイト）	1-20.Webサービス	20.外部リンク	
		20.バンクの場合、収集対象		21.国	1-21.ユーザ登録	21.論文等（PubMed ID）	
				22.提供機関	2-1.データ名		
				23.提供機関（略称）	2-2.データ内容の説明		
				24.提供機関ID	2-3.データファイル		
				25.DB開始日	2-4.データ取得方法		
				26.DB更新日	2-5.解析方法		
				27.横断検索	2-6.データ件数		
				28.DBアーカイブ			
				29.プロジェクト一覧			
				30.メタデータ更新日			
				31.メタデータ作成者、更新者			

＊：独立の項目立てではない。

新カタログ項目案

#	項目名	説明	備考
基本情報			
1	ID	NBDC共通のデータベースID	NBDCで採番
2	名称	データベースの正式名称	
3	略称	データベースの略称	略称がなければ、データベース名と同一にする
4	URL	データベースをWebで公開している場合、そのURL	
5	言語	データベースを記述する言語	ja, enなどISO 639の２文字表記で表す複数の言語で運用されている場合は、スラッシュ区切り（ja/enなど）で表す
6	運用機関名	現在データベースを運用している機関名	J-GLOBAL収録の研究機関名で表記
7	運用機関所在国	運用機関が所在する国名	国名は、ISO 3166-1の英語名で表記
8	上位データベース	当該データベースが他のデータベースの一部になっている場合は、そのデータベース名を記載	
9	生物種	データベースが扱う生物の種名	扱う生物種が少なければ、そのTaxonomy IDとNameを記載する 生物種が多い場合は、それらを一括した呼び名で記載しても良い
10	データの種類	データベースに登録されたデータの種類（ゲノム配列、マイクロアレイデータ等）	タグ形式で、１つのデータベースに複数付与可能 タグはNBDCが管理し、別途管理規則を定める
11	その他のキーワード	生物種、データの種類以外のキーワード（データベースの表示方法等）	タグ形式で、１つのデータベースに複数付与可能 タグはNBDCが管理し、別途管理規則を定める
12	説明	データベースの説明	
13	論文等	当該データベースに深く関連した論文のPubmedID	
14	最終更新日時	データベースが最終的に更新された日時	データベースサイトに自動的にアクセスして情報を収集する
拡張情報（データベース自体ではなく、このカタログの利便性向上のための情報）			
1	データベースカタログ（旧）へのリンク	生命科学系データベースカタログ（旧）のTogoDBサイトへのリンク	
2	WingPro（旧）へのリンク	WingPro（旧）のTogoDBサイトへのリンク	
3	アーカイブへのリンク	生命科学系データベースアーカイブへのリンク	
4	MEDALSへのリンク	MEDALSへのリンク	

＜目的＞
ライフサイエンス分野に存在する可能な限り**多くのデータベースの所在情報の提供**

＜項目設定基準＞
➢DBの概要を理解できる必要最低限の項目を選定
➢記載項目を絞り、情報収集作業時間を低減
➢データセットの利用／交換のためのメタデータはLSDBアーカイブ（ポスター番号2参照）から提供

今後：コンテンツの充実と新サイト構築

●新メタデータ項目に基づいたカタログ内容整備
●統一IDを用いてLSDBアーカイブ等他サービスへのリンクづけ
●新サイト構築

各省のデータベース統合事業との連携

各省のデータベース統合関連事業と連携し、カタログ情報の集約、更新を進める。

MEDALS（経済産業省・産業技術総合研究所）
http://medals.jp/
AgriTOGO（農林水産省・農業生物資源研究所）
http:/togo.dna.affrc.go.jp/
厚生労働省・医薬基盤研究所
http://www.nibio.go.jp/

各事業所管のデータベース情報を受領し、NBDCのカタログデータに取り込む。

今後：連携の深化

●各省のカタログサービスとの相互リンクを追加
●各省とのデータベース情報交換の効率化により更新作業の改善を図る

トーゴーの日シンポジウム 2011　ライフサイエンス分野のデータベース統合の“カタチ”を探る

会期：2011.10.5　於：日本科学未来館　みらい CAN ホール

Licensed under a Creative Commons表示2.1日本 license (c)2011 バイオサイエンスデータベースセンター