

生命科学系データベースアーカイブ
作成ガイドライン

第 3.2 版

バイオサイエンスデータベースセンター

目次

1	はじめに	4
2	改訂履歴	5
3	生命科学系データベースアーカイブの目的・用途	6
3.1	アーカイブ受け入れ方針	6
3.1.1	アーカイブの受け入れ対象のデータベース	6
3.1.2	アーカイブの受け入れ対象外のデータベース	6
3.2	アーカイブ利用者のメリット	7
3.2.1	アーカイブ利用者のメリット	7
3.2.2	DB 寄託者のメリット	7
4	アーカイブ作成の概要	8
5	アーカイブファイルの作成	10
5.1	アーカイブファイル仕様検討	10
5.1.1	標準的なファイル構成	10
5.1.2	ファイル構成の設計	11
5.1.3	データ項目の決定	13
5.1.4	データ変換方法の検討	14
5.2	アーカイブファイルの作成	15
5.2.1	文字コード別ファイルの作成	15
5.2.2	アーカイブファイルの命名規則	15
5.2.3	アーカイブファイルのディレクトリ構造	15
5.2.4	アーカイブファイルのチェック	16
5.2.5	アーカイブファイルの圧縮	16
5.3	アーカイブファイル送付	16
6	簡易検索サイトの作成（オプション）	17
6.1	アカウントの取得	17
6.2	アーカイブファイルのアップロード	17
6.3	簡易検索サイトのレイアウト調整	17
6.4	簡易検索サイトの公開	17
7	メタデータの作成	19
7.1	データベースのメタデータ	20
7.2	アーカイブファイルのメタデータ	25
7.3	NAR カテゴリリスト	26
8	利用許諾決定	30

8.1	利用許諾テンプレート	30
8.2	標準利用許諾について	32
8.2.1	クリエイティブ・コモンズ・ジャパンについて.....	33
8.2.2	利用者に許可されていること	33
8.2.3	利用者が守るべき条件.....	33
8.2.4	データベースの営利利用及び改変について	33
8.3	標準利用許諾を利用しない場合	35
8.4	利用許諾の入力.....	35
9	お問い合わせ先	36

1 はじめに

本ガイドラインは、文部科学省・経済産業省・厚生労働省・農林水産省（以下、連携 4 省）が連携して進めているライフサイエンスデータベースの統合の一環として、各研究機関が作成・保有している生命科学系のデータベースをバイオサイエンスデータベースセンター（以下、NBDC）からアーカイブとして公開するにあたり、実施すべき手順を記載したものです。

本ガイドラインには、データベースのアーカイブ化に必要な作業—アーカイブの作成、メタデータの作成、利用許諾の決定等—について標準的な手順をまとめています。各研究機関におかれましては、これらの手順に従ってアーカイブを作成していただきますようお願いいたします。

2 改訂履歴

版番号	作成・改訂日	改訂理由
第1版	2010/11/30	新規作成
第2版	2011/4/1	<ul style="list-style-type: none">・ガイドライン発行組織の変更・利用許諾決定プロセスの変更・お問い合わせ先の変更
第3版	2012/5/14	<ul style="list-style-type: none">・「アーカイブ受け入れ方針」の追加・アーカイブファイル作成の変更・英語版のメタデータ、利用許諾に関する記述を追加・その他記述内容の修正
第3.1版	2015/4/1	<ul style="list-style-type: none">・お問い合わせ先の情報の更新
第3.2版	2015/11/17	<ul style="list-style-type: none">・利用許諾テンプレートの変更を反映

3 生命科学系データベースアーカイブの目的・用途

生命科学系データベースアーカイブ（以下、アーカイブ）は、国内の生命科学研究者が生み出したデータセットをわが国の公共財としてまとめて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するサービスです。

3.1 アーカイブ受け入れ方針

アーカイブでは、以下のデータベース受け入れ対象とします。アーカイブでは、データベース寄託を広く受け入れています。

3.1.1 アーカイブの受け入れ対象のデータベース

アーカイブでは、以下の条件に当てはまるデータベースまたはデータを受け入れ対象とします。

- 生命科学分野のデータベース
- アーカイブの仕組みを利用してデータ公開を希望するデータ

なお、更新が頻繁でアーカイブにおけるアーカイブデータと同期が取りづらいなどの理由でアーカイブの Web サイト上での公開は不都合であると判断された場合は、アーカイブでメタデータを公開し、アーカイブデータはオリジナルの Web サイトまたは FTP サイトで公開されたアーカイブデータにリンクするなどの対応を行うことができます。このような場合は、NBDC データベース寄託担当者にご相談ください。

3.1.2 アーカイブの受け入れ対象外のデータベース

以下のケースについては、基本的にアーカイブでのデータ受け入れ対象外としますが、対応可能な場合もありますので、アーカイブ担当者にご相談ください。

- データベース作成者・管理者と連絡が取れないデータベース
アーカイブデータ公開時の利用許諾の承認の都合上、必要です。
- 個人情報を含むデータベース。
データベース寄託元で公開の了解を得られた個人情報については協議を行い、受け入れを検討します。
- データが大量で、アーカイブで受け入れることが困難なデータベース。

合計 1TB を超えるデータの寄託をご検討される場合はご相談ください。

3.2 アーカイブ利用者のメリット

アーカイブの利用者、及びアーカイブにデータベースを寄託していただく各研究機関（以下、DB 寄託者）にとってのアーカイブのメリットは以下の通りです。

3.2.1 アーカイブ利用者のメリット

1. 大部分のデータがダウンロード可能ですので、データマイニングや統合データベース構築に利用いただくことが可能です。全レコードを一括してダウンロードできるだけでなく、簡易検索によってレコードを絞り込んでダウンロードすることもできます（絞り込めないデータもあります）。
2. 詳細なメタデータのご提供により、必要なデータへのアクセスと利用が非常に容易になります。また、利用許諾条件が明示されていますので、安心してデータを利用することができます。

3.2.2 DB 寄託者のメリット

1. DB 寄託者が望む限り、データベースはアーカイブとして保全されます。
2. データをアーカイブに集約することにより、より広い分野の利用者からのアクセス増大を期待できます。
3. 標準利用許諾を用意していますので、法律の知識に詳しくなくても、手間をかけずに利用許諾を設定することができます。
4. アーカイブ化したデータベースが検索で閲覧されるだけでなく、ダウンロードされてデータマイニングや統合データベース構築等により深く利用されることにより、生命科学や医学、医療などによりいっそうの貢献をすることができます。

4 アーカイブ作成の概要

アーカイブの作成フローは下図の通りです。なお、図中の「DB保有者」はオリジナルデータベースを作成した研究機関等を指し、「各省取りまとめ機関」は連携4省のそれぞれにおいて、所管する研究機関等とNBDCとの間の窓口となる機関を指します。

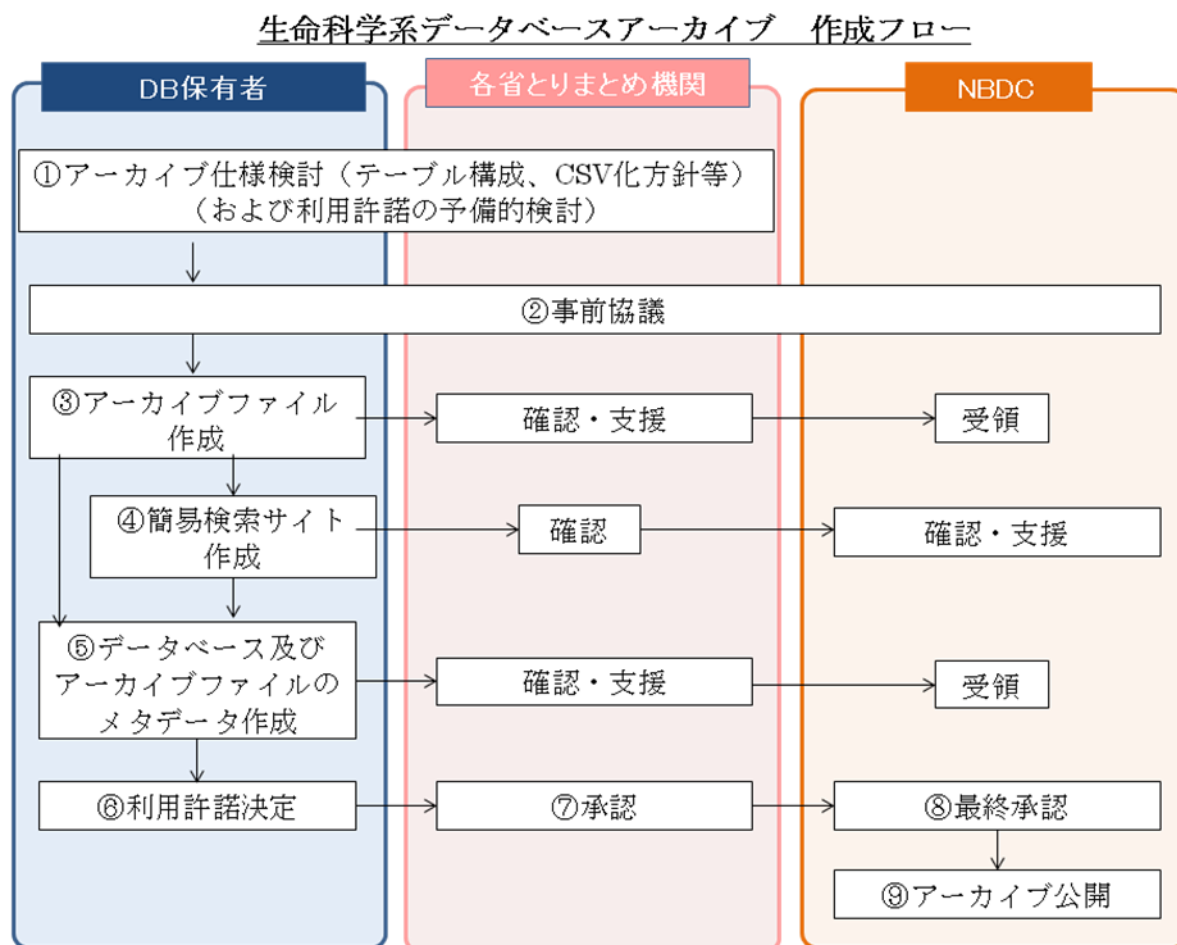


図 4-1 生命科学系データベースアーカイブ 作成フロー

フロー中の各工程の概要は下表の通りです。

表 4-1 生命科学系データベースアーカイブ 作成フローの各工程

#	工程	概要	関連章・節
①	アーカイブ仕様検討	アーカイブファイルの構成やデータ項目等を設計し、オリジナルデータをこの構成に変換する方法を検討します。	5.1
②	事前協議	データの内容、規模、公開時期等について	

		て DB 保有者と各省とりまとめ機関・NBDC との間で協議します。	
③	アーカイブファイル作成	オリジナルデータを、①で設計したアーカイブファイルに変換します。	5.2
④	簡易検索サイト作成 (オプション)	アーカイブファイルを TogoDB (Web データベースシステム) にアップロードして簡易検索サイトを作成します。この工程はオプションであり、必須ではありません。	6
⑤	データベース及びアーカイブファイルのメタデータ作成	データベースの内容及びアーカイブファイルの各項目に関する詳細情報 (メタデータ) を作成します。	7
⑥	利用許諾決定	データベースをアーカイブで公開するにあたっての利用許諾を決定します。	8
⑦	承認	各省とりまとめ機関において、メタデータの内容と利用許諾について確認し、アーカイブ公開することを承認します。	
⑧	最終承認	NBDC において、アーカイブデータ・メタデータ・利用許諾の文言等を確認し、アーカイブ公開することを承認します。	
⑨	アーカイブ公開	NBDC よりアーカイブを公開します。	

5 アーカイブファイルの作成

ここでは、アーカイブ仕様検討からアーカイブファイル送付までについて説明します。

5.1 アーカイブファイル仕様検討

NBDC が想定するユーザのファイル利用法は以下の通りです。

- Excel などのアプリケーションソフトを使ったデータ内容の閲覧
- ユーザ独自のプログラムを使ったデータの処理・加工
- 簡易検索サイトでのデータの検索・閲覧

アーカイブファイルの仕様を検討するにあたっては、こうした用途を考慮する必要があります。

5.1.1 標準的なファイル構成

アーカイブファイルの標準的なファイル構成は下図の通りです。

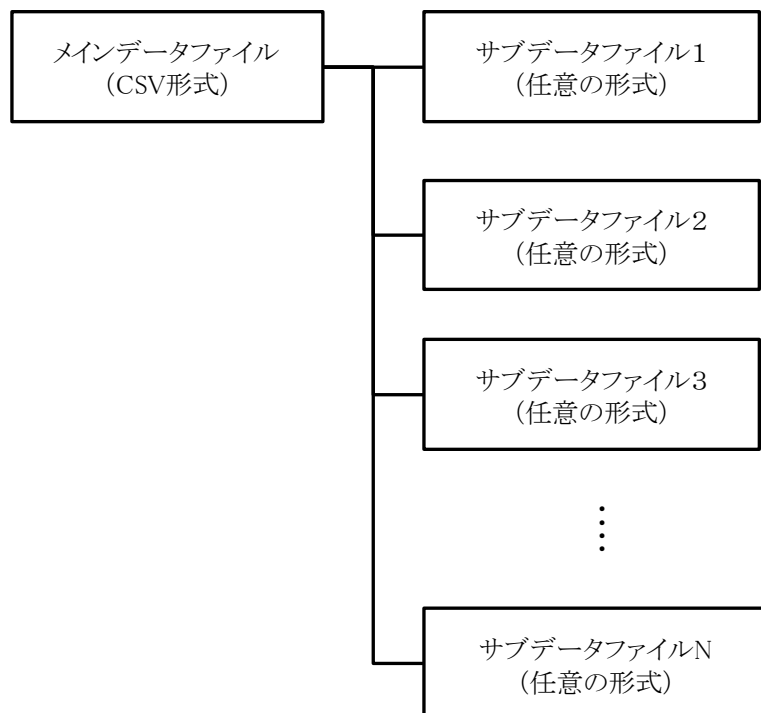


図 5-1 標準的なアーカイブファイルの構成

- **メインデータファイル：**

アーカイブのメインデータを格納するファイルです。1つのアーカイブに必ず1つ作成します。このファイルでは、データをリスト化して CSV 形式で表します。オリジナルデー

データベースの内容がテキストデータのみであり、しかも1つのテーブルのみから構成されている場合、アーカイブファイルはメインデータファイルのみとなります。

サブデータファイルがある場合は、メインデータとサブデータ間で関連付けられるように、項目を設定します。

- サブデータファイル：

メインデータファイルから派生する CSV 形式データの場合、オリジナルデータベースの内容にテキスト以外のファイル（画像など）が含まれている場合、テキストデータであっても XML フォーマットなど CSV 形式では表現できないものが含まれている場合、これらのデータを、メインデータファイルとは別にサブデータファイルとします。

例えば、画像と画像の説明情報（テキスト）から構成されるデータベースをアーカイブ化する場合を想定します。まず、画像の説明情報からメインデータファイルを作成し、画像ファイルをサブデータファイルとします。このとき、メインデータファイルには画像ファイルを一意に特定できる情報（ファイル名など）を必ず入力しなければなりません。

※メインデータファイル、サブデータファイルの標準的なデータ構成で設計できない場合がある場合は、NBDC アーカイブ担当者にご相談ください。

5.1.2 ファイル構成の設計

まず、アーカイブファイルの構成を設計します。オリジナルデータベースのどのテーブル内容をメインデータファイルとするか、何個のサブデータファイルを作成するか等を検討する必要があります。

5.1.2.1 メインデータファイルとサブデータファイルの設計ガイドライン

データを一覧するための目次の役割のメインデータを設定し、メインテーブルから派生するデータがある場合は、サブセットのデータ（サブデータファイル）を定義します。サブデータは、表形式データに限らず、画像ファイルやドキュメントファイルも含まれます。データ構造によっては、サブデータをさらに階層化することが有り得ますが、最低限の階層になるよう設計してください。

ファイル構成を決定する際に、ユーザがデータを一覧でき、データの内容を解釈しやすいように、データを統合してデータファイル数が少なくなるように設計します。

メインデータファイルの設計は、以下のガイドラインに沿って設計してください。

- CSV 形式の表形式のデータを作成してください。
- メインデータは、データベースの入り口の役割となることを意識し、データベースの中心となるデータのエントリーID と、エントリーの情報を含むようにしてください。

- サブデータが存在する場合は、メインデータとサブデータ間で参照できるように外部キーとなる項目を設定してください。

5.1.2.2 サブデータファイルの種類

メインデータから派生するサブデータとなるファイルは、複数の種類が考えられます。以下は、ファイル種類別にサブデータファイルを作成する際の注意事項です。

- 表形式データ
 - ファイル形式は CSV 形式とします。
 - 改行を含むテキストも許します。
 - サブデータが存在する場合は、メインデータとサブデータの間で参照できるように外部キーとなる項目を設定してください。
- Excel データ(.xls, .xlsx 形式)
 - Microsoft Excel で作成したファイルは、Excel をユーザが所持していないと閲覧、編集できないため、CSV 形式に変換してアーカイブファイルを作成してください。
 - 但し、Excel 形式のファイルが、論文のサプリメントデータとなっている場合や、オリジナルサイトでその形式で公開していた場合は、ファイルでの表現を保持するために、そのままでも構いません。
 - Microsoft Office 2007 以降の Excel で作成された .xlsx 形式ファイルの場合は、Office 2003 以前のユーザが編集、閲覧できるように、xls 形式ファイルもアーカイブファイルに追加してください。
- 画像データ
 - ファイル名をメインデータ、サブデータ内で使用される ID を使用するなどして、どのデータと関連があるかを明確にしてください。
 - 画像フォーマットは Web ブラウザで閲覧できる形式 (JPEG、PNG、GIF、TIFF 等) を推奨します。画像の閲覧に特定のソフトウェアが必要な場合は、その旨をメタデータに明記してください。
 - 画像データは、他のユーザが高精度に閲覧できるように、高解像度のファイルを推奨します。
 - ◇ 後述する簡易検索サイト作成用にはウェブブラウザでの閲覧用に低解像度のファイルを作成してください。
- 国際標準フォーマット／実験機器の出力データ／解析ツールによる解析結果
 - ファイル名をメインデータ、サブデータ内で使用される ID を使用するなどして、どのデータと関連があるかを明確にしてください。
 - FASTA 形式や GFF 形式などライフサイエンス分野で標準的なファイルフォーマットの場合はそのままの形式で構いません。簡易検索サイトでの閲覧をする場合は、CSV 形式に変換します。

- 相同性検索結果などの解析ツールによる解析結果で、CSV ファイルに格納するにはサイズが大きいと判断される場合は、結果をテキストファイルにし、サブデータとしてください。それらのテキストファイルのファイル名を、関連するデータの ID を使用するなどして、メインデータまたはサブデータの表形式データと関連付けを明確にしてください。
- ドキュメントファイル（補足資料、論文、報告書、など）
 - ファイル名をメインデータ、サブデータ内で使用される ID を使用するなどして、どのデータと関連があるかを明確にしてください。
 - 論文、報告書などのデータファイルは、PDF 形式、HTML 形式を推奨します。

5.1.2.3 他のデータベース由来のデータについて

- 参照したデータベースのバージョンまたは日付をメタデータ内に記述してください。
- データ内に、他のデータベース由来のデータを含む場合は、参照したデータベースの利用許諾を確認してください。参照したデータベースの利用許諾の条件に、データの再配布の禁止や、利用許諾の継承が含まれる場合は、各省の取りまとめ機関または NBDC と対応を協議してください。

5.1.3 データ項目の決定

次に、データ項目を決定します。オリジナルデータベースのデータ項目を、以下の観点から再度検討してください。

5.1.3.1 データ中の ID 項目について

- 主キーとなる項目について
 - 1 エントリーが一意にきまる主キー、または複合主キーを設定してください。
 - オリジナルのデータで主キーが明確でない場合、連番(1,2,3…)の ID を新たに割り振り、主キーとしてください。
 - 主キーについて、メタデータに記述してください。
- 外部キーとなる項目について
 - 外部キーを設定する場合、対応関係を明確にするために参照先テーブルのカラム名と揃えることを推奨します。
- 他の公共データベースへの参照
 - 他のデータベース（GenBank、UniProt、SwissProt 等）のエントリーと関連付けられる場合、そのデータベースの ID を含むようにしてください。
 - 複数の ID を記載する項目には、ID で使用される文字列を考慮し、適切な区切り文字で区切って表記してください。

5.1.3.2 項目名の命名規則

- データベースのオリジナルサイトでの項目名と一致するように命名することを推奨します。一致しない場合は、メタデータにオリジナルの項目と対応を記述してください。
- 項目名の重複が起きないように命名してください。
- 他のデータベースを参照する場合、そのデータベース名を項目名に明示するか、データのメタデータに参照先データベースの情報を記述してください。公共 DB の ID を含む項目名は、「(データベース名) + ID」または「(データベース名) + Accession」のように命名してください。

例： UniProt ID の場合 → UniProt ID
GenBank の場合 → GenBank Accession

5.1.3.3 データ型の統一

- String (テキスト)、Text (可変長テキスト)、Int (整数)、Float (少数) などのデータ型を想定し、できるだけ項目内に同じ型の値で表現できるようにしてください。
- 有効数字を設定する場合、桁が揃っているかを確認してください。
- データ型を統一した場合は、データの説明 (メタデータ) に記述してください。

5.1.3.4 項目の統合・分割

- 上記の諸条件を考慮して、必要に応じて項目を統合・分割してください。
- データの一覧性を重視し、データ項目を細分化しすぎないようにしてください。
- 1つの項目に複数の値が入る場合がありますが、区切り文字を使用して区切ってください。

5.1.3.5 簡易検索サイト用の項目

- TogoDB を利用した簡易検索サイトでは、HTML タグを利用した画像やハイパーテキストの表示が可能です。簡易検索用サイトでそのような機能を利用したい場合は、HTML タグを含む簡易検索用サイト用の項目、データを用意してください。
- 項目の値全体にリンクを付与する場合は、TogoDB の設定で付与することが可能ですので、その場合は TogoDB 用の項目、データの追加は必要ありません。

5.1.4 データ変換方法の検討

オリジナルデータを 5.1.2 で決定したアーカイブデータに変換する方法を検討してください。Perl、PHP などのスクリプト言語で変換プログラムを作成するのが一般的です。

5.2 アーカイブファイルの作成

5.1 で決定したアーカイブの仕様と変換方法に基づいてアーカイブファイルを作成します。

5.2.1 文字コード別ファイルの作成

メインデータファイル、及びサブデータファイルの中でも CSV 形式のファイルは、ファイル内に 2 バイト文字（日本語など）を含む場合には、Shift-JIS 及び UTF8 の 2 つの文字コードを使ってそれぞれ作成してください。Shift-JIS で書き込まれたファイルは、Excel で閲覧する場合に利用します。UTF8 で書き込まれたファイルは、プログラムで処理する場合に便利です。

5.2.2 アーカイブファイルの命名規則

- ファイル名に利用できる文字は、英数字及び”_”（アンダーバー）のみとします。

5.2.2.1 データの内容が英数字のみのアーカイブファイルの命名規則

- メインデータファイルの名前は、”XXXXX_main.csv”とします。（XXXXX は任意の英数字からなる文字列）
- サブデータファイルのうち CSV 形式のファイルの名前は、”XXXXX.csv”とします。（XXXXX は任意の英数字からなる文字列）

5.2.2.2 データに 2 バイト文字（日本語など）を含むアーカイブファイルの命名規則

- メインデータファイルの名前は、”XXXXX_main_sjis.csv”（Shift-JIS 用ファイル）及び”XXXXX_main_utf8.csv”（UTF8 用ファイル）とします。（XXXXX は任意の英数字からなる文字列）
- サブデータファイルのうち CSV 形式のファイルの名前は、”XXXXX_sjis.csv”（Shift-JIS 用ファイル）及び”XXXXX_utf8.csv”（UTF8 用ファイル）とします。（XXXXX は任意の英数字からなる文字列）

5.2.2.3 CSV 形式以外のアーカイブファイルの命名規則

- ファイル名にデータエントリの ID 名を含むなどして、メイン/サブデータのエントリーとの対応関係がわかるようにしてください。
- ファイル名にフォーマットを判別できる拡張子を付けてください。

5.2.3 アーカイブファイルのディレクトリ構造

- 画像ファイルなどの同種の内容のファイルは、ディレクトリ（フォルダ）に分けて整理して保存してください。

5.2.4 アーカイブファイルのチェック

- 誤字、脱字、スペルミスのチェックをしてください。特に日本語の場合、同じ意味で使用される文字列が、全角、半角で混在することがありますので、表現を統一してください。
- 表形式のデータの編集に Microsoft Excel を使用した場合、データ読み込み時に値を自動的に変換される場合がありますので、オリジナルの値が損なわれないように注意してください。

5.2.5 アーカイブファイルの圧縮

作成したアーカイブファイルは、ZIP 形式で圧縮してください。同一種類のサブデータファイルの数が多いものは、1つのフォルダにまとめてから圧縮してください。

5.3 アーカイブファイル送付

作成したアーカイブファイルは、NBDC 宛（アドレスは 9.お問い合わせ先 参照）にメールに添付して送付してください。

ただし、送付するファイルの総量が 10MB を超える場合は NBDC までご相談ください。

6 簡易検索サイトの作成（オプション）

アーカイブでは、アーカイブファイルの他にオプションとして簡易検索サイトを利用者に提供しています。簡易検索サイトでは、主にメインデータファイル及び CSV 形式のサブデータファイルを検索・参照し、必要なデータだけを部分的にダウンロードすることが可能です。

簡易検索サイトは、ライフサイエンス統合データベースプロジェクトにおいて開発した Web データベースシステム TogoDB を利用しています。簡易検索サイトを作成するには、アーカイブファイルを TogoDB 上にアップロードする必要があります。

なお、簡易検索サイトの作成は必須ではありませんが、利用者の利便性を高めるためになるべく簡易検索サイトを作成していただくようお願いします。

6.1 アカウントの取得

TogoDB を利用するには DBCLS OpenID アカウントが必要です。OpenID アカウントの取得は、DBCLS OpenID サービス (<http://openid.dbcls.jp/>) にて行うことができます。

6.2 アーカイブファイルのアップロード

アーカイブファイルを TogoDB にアップロードします。アップロードできるファイルは、CSV 形式のみです。それ以外の形式のファイルは TogoDB にアップロードすることができません。アップロード方法の詳細は、「TogoDB 利用者向け操作説明書」 (http://dbarchive.biosciencedbc.jp/files/togodb_user.pdf) を参照してください。

TogoDB 上で画像データの表示を行う場合、ファイルを NBDC の Web サーバにアップロードする必要がありますので、その場合は別途 NBDC アーカイブ担当までご相談ください。

6.3 簡易検索サイトのレイアウト調整

簡易検索サイトには、独自のヘッダ・フッタをつけることができます。詳細については、「TogoDB 利用者向け操作説明書」を参照してください。NBDC にてヘッダのテンプレートを用意していますので、ご利用ください。

6.4 簡易検索サイトの公開

簡易検索サイトの作成が終わりましたら、NBDC でサイト内容の確認を行います。サイトの URL を NBDC までご連絡ください (9.お問い合わせ先 参照)。

NBDC での確認が終わった後、サイトの公開を行ってください。サイトの公開方法は

「TogoDB 利用者向け操作説明書」を参照してください。

7 メタデータの作成

アーカイブファイルの作成後、データベース及びアーカイブファイルについての詳細情報（メタデータ）を記入します。メタデータの記入は、NBDCが指定したWebフォームに直接入力する形で行います。

日本語版、英語版のメタデータフォーマットがありますが、先に日本語版のメタデータを公開し、その後英語版のメタデータを公開することができます。

Webフォームへの入力方法等については、NBDCから別途ご案内します。ここでは、メタデータの各項目の意味と記入例を示します。

7.1 データベースのメタデータ

データベース全体に対してメタデータを作成します。メタデータの各項目の意味と記入例は下表の通りです。

表 7-1 データベースメタデータ

#	項目(日本語版)	項目(英語版)	記入例	説明	
データベース全般 (General information of database)					
1	名称	Database name	DIAM	データベースの名称です。“DDBJ”など短い名前とします。短い名前がなく、長い名前のみ存在する場合は、それを名称としてください。	
2	読み方	-	ダイヤモンド	名称の正しい読み方をカタカナで記入してください。	
3	別名	Alternative name	重原子データベース	“DNA Data Bank of Japan”など長い名前とします。	
4	作成者	Creator(s)	作成者氏名	*統合 太郎	主にデータベースの作成に関わった方の氏名と所属です。複数記入することができます。代表者には“*”をつけてください。
			英名	Taro Tougou	
			所属	バイオサイエンスデータベースセンター	
			作成者氏名	統合 花子	
			英名	Hanako Tougou	

			所属	バイオサイエンスデータベースセンター	
5	連絡先	Contact address	〒102-0081 東京都千代田区四番町5-3 サイエンスプラザ TEL 03-5214-8491 FAX 03-5214-8470		データベースに関する問い合わせ先です。メールアドレス・住所・電話番号・担当者名などを記入してください。
6	データベース分類	Database classification	器官データベース		該当するものをNARカテゴリリスト(表7-3)から選択します。複数選択が可能です。NARカテゴリと、データベースの分類の例は、 http://www3.oup.co.uk/nar/database/cap/ を参照してください。「その他(自由記述)」を選択した場合は、具体的に記入してください。
7	生物種 (Taxonomy Name/Taxonomy ID)	Organism (Taxonomy Name/Taxonomy ID)	Homo Sapiens (Taxonomy ID: 9606)		データベースが対象としている生物種の Taxonomy ID と Taxonomy Name (NCBI(http://www.ncbi.nlm.nih.gov/sites/entrez?db=taxonomy)または DDBJ(http://txsearch.ddbj.nig.ac.jp/top-j.html)のサイトで参照できます)を記入してください。複数生物種の記入が可能です。生物種が多すぎる場合は、Taxonomy Name の記入欄に、対象生物種を参照できる資料名・URL 等を記入してください。
8	説明	Database description	解剖学用語が示す人体の部品(臓器、器官)の位置と形状を3次元人体モデルで記述したデータベース		データベース内容の説明です。

9	データベースの特長・有用性・活用方法	Features and manner of utilization of database	人体3次元モデルが無償利用できるのが最大の特徴。臓器の3次元的位置や形状を確認する電子アトラスとしてだけでなく、人体シミュレーションやデータマッピングなどの入力データとして再利用可能である。		現在考えられるデータベースの特長、有用性、活用方法です。データの特長・データの活用方法など、データベースをPRする内容を記入してください。
10	予算的背景・プロジェクト	Background and funding	文部科学省「ライフサイエンス統合データベース」プロジェクト		データベース作成を主に資金面で支援した制度やプロジェクトについて記入してください。
11	論文等	Reference(s)	文献名 (Article title)	BodyParts3D: 3D structure database for anatomical concepts.	データベースの作成事実や内容を発表した論文などの文献があれば、文献名・著者名・発表年・所収雑誌名などを記入してください。
			著者名 (Author name(s))	Mitsuhashi N, Fujieda K, Tamura T, Kawamoto S, Takagi T, Okubo K.	
			雑誌名/掲載年月/号 (Journal title / Publication date / Issue)	Nucleic Acids Res. 2008 Oct 3.	
			Pubmed ID	18835852	
データベースのオリジナルサイト情報 (Original website information)					
12	データベース運用場所	Database maintenance site:	ライフサイエンス統合データベースセンター		オリジナルサイトが運用されている場所(機関など)です。
13	オリジナルサ	URL of the	http://lifesciencedb.jp/bp3d//		データベースのオリジナルサイトの URL です。

	イト	original website		
14	運用開始年月日	Operation start date	2007/10/1	データベースの運用が始まった日付をわかる範囲で記入してください。
15	最終更新年月日	Last updated date	2009/2/1	最後にデータベースが更新された(単にページが更新されただけでなく、データ自体が更新された)日付をわかる範囲で記入してください。
16	統括サイト	URL of the portal site	http://lifesciencedb.jp/	データベースのオリジナルサイトを統括するポータルサイトなどの URL を記入してください。
17	データの一括ダウンロードサイト	Whole data download	http://lifesciencedb.jp/bp3d/info/download/index.html	オリジナルサイトにおいて、ユーザがデータベースのデータを HTTP または FTP 経由で一括してダウンロードできる場合には、ダウンロードページの URL を記入してください。
18	参照先データベース	Referenced database	Foundational Model of Anatomy	データベース中の各データからリンクされているデータベース(例えば説明図3のようなもの)を記入してください。データベースが複数ある場合は、全てのデータベースを記入してください。
19	一覧表示	Entry list	有り	オリジナルサイトにおいて、データが一覧になっている場合は「有り」を、そうでなければ「無し」を選択してください。
20	クエリ検索	Query search	無し	オリジナルサイトにおいて、検索クエリを入力してデータを検索できる場合は「有り」を、そうでなければ「無し」を選択してください。
21	Web サービス	Web services	有り	オリジナルサイトにおいて、SOAP や REST 等の Web サービスを提供している場合は「有り」を、そうでなければ「無し」を
22	Web サービス	URL of Web	http://lifesciencedb.jp/ag/api/index.jsp	

	URL	services:		選択してください。また、Web サービスについての説明ページがある場合は、その URL を記入してください。
23	ユーザ登録	Need for user registration:	無し	オリジナルサイトにおいて、データベース中のデータにアクセスするためにユーザ登録が必要であれば「有り」を、そうでなければ「無し」を選択してください。

7.2 アーカイブファイルのメタデータ

アーカイブファイルごとに1つのメタデータを作成します。メタデータの各項目の意味と記入例は下表の通りです。

表 7-2 アーカイブファイルメタデータ

#	項目（日本語版）	項目（英語版）	記入例	説明
1	データ名	Data name	安全性に関する文献	データの名称です。
2	データ内容の説明	Description of data contents	バイオテクノロジーの安全性に関する文献の書誌情報などの情報	データ内容の説明です。
3	データファイル	Data file	diam_safety_literature.zip (84.7MB)	(NBDC にて作成)データファイルへのリンクです。
4	データ取得方法	Data acquisition method	文献	データの取得元です。実験・公共DB・文献 から該当するものを選択します。複数選択することができます。
5	解析方法	Data analysis method	検索	データの取得方法です。データソースが実験の場合は装置や実験方法を記入します。データソースが公共DBの場合は、どのDB から取得したかを記入します。データソースが文献の場合は、文献ソース名と取得方法等を記入します。複数記入することができます。
6	データ件数	Number of data entries	<ul style="list-style-type: none"> ・1970年代の文献:2件 ・1980年代の文献:12件 ・1990年代の文献:27件 ・2000年代の文献:47件 	データの件数です。

7.3 NAR カテゴリリスト

データベースのメタデータ作成の際に参照する、NAR (Nucleic Acids Research)による生命科学系データベースのカテゴリリストです。詳細は、

<http://www3.oup.co.uk/nar/database/cap/> を参照してください。

表 7-3 NAR カテゴリリスト (日本語版)

#	NAR カテゴリ
1	塩基配列データベース
2	塩基配列データベース-国際塩基配列データベース連携
3	塩基配列データベース-コーディング/ノンコーディング領域 DNA
4	塩基配列データベース-遺伝子構造、イントロン/エクソン、スプライス部位
5	塩基配列データベース-転写調節部位、転写因子
6	RNA 配列データベース
7	タンパク質配列データベース
8	タンパク質配列データベース-配列データベース全般
9	タンパク質配列データベース-タンパク質属性
10	タンパク質配列データベース-タンパク質の局在とターゲティング
11	タンパク質配列データベース-タンパク質配列モチーフ、タンパク質活性部位
12	タンパク質配列データベース-タンパク質ドメインデータベース、タンパク質分類
13	タンパク質配列データベース-タンパク質ファミリー別データベース
14	構造データベース
15	構造データベース-低分子
16	構造データベース-炭水化物(Carbohydrates)
17	構造データベース-核酸構造
18	構造データベース-タンパク質構造
19	ゲノクスデータベース(脊椎動物以外)
20	ゲノクスデータベース(脊椎動物以外)-ゲノムアノテーション用語、オントロジー、命名法
21	ゲノクスデータベース(脊椎動物以外)-生物分類/同定
22	ゲノクスデータベース(脊椎動物以外)-ゲノクスデータベース全般
23	ゲノクスデータベース(脊椎動物以外)-ウィルスゲノムデータベース
24	ゲノクスデータベース(脊椎動物以外)-原核生物ゲノムデータベース
25	ゲノクスデータベース(脊椎動物以外)-単細胞真核生物ゲノムデータベース
26	ゲノクスデータベース(脊椎動物以外)-真菌ゲノムデータベース

27	ゲノクスデータベース(脊椎動物以外)-無脊椎動物ゲノムデータベース
28	代謝系/シグナル伝達経路
29	代謝系/シグナル伝達経路-酵素、酵素命名法
30	代謝系/シグナル伝達経路-代謝経路
31	代謝系/シグナル伝達経路-タンパク質・タンパク質相互作用
32	代謝系/シグナル伝達経路-シグナル伝達経路
33	ヒト/その他の脊椎動物ゲノム
34	ヒト/その他の脊椎動物ゲノム-モデル生物、比較ゲノミクス
35	ヒト/その他の脊椎動物ゲノム-ヒトゲノムデータベース/マップ/ビューワ
36	ヒト/その他の脊椎動物ゲノム-ヒト ORF
37	ヒト遺伝子/疾患
38	ヒト遺伝子/疾患-人類遺伝学データベース全般
39	ヒト遺伝子/疾患-多型データベース全般
40	ヒト遺伝子/疾患-ガン遺伝子データベース
41	ヒト遺伝子/疾患-特定の遺伝子/器官系/疾患に関するデータベース
42	マイクロアレイデータ、その他の発現データのデータベース
43	プロテオミクス関連
44	その他の分子生物学データベース
45	その他の分子生物学データベース-薬剤、ドラッグデザイン
46	その他の分子生物学データベース-分子プローブ、プライマー
47	オルガネラデータベース
48	オルガネラデータベース-ミトコンドリア遺伝子/タンパク質
49	植物データベース
50	植物データベース-植物データベース全般
51	植物データベース-シロイヌナズナ
52	植物データベース-イネ
53	植物データベース-その他の植物
54	免疫学データベース
55	その他(自由記述)

表 7-4 NAR カテゴリリスト (英語版)

#	NAR Categories List
1	Nucleotide Sequence Databases

2	Nucleotide Sequence Databases - International Nucleotide Sequence Database Collaboration
3	Nucleotide Sequence Databases - Coding and non-coding DNA
4	Nucleotide Sequence Databases - Gene structure, introns and exons, splice sites
5	Nucleotide Sequence Databases - Transcriptional regulator sites and transcription factors
6	RNA sequence databases
7	Protein sequence databases
8	Protein sequence databases - General sequence databases
9	Protein sequence databases - Protein properties
10	Protein sequence databases - Protein localization and targeting
11	Protein sequence databases - Protein sequence motifs and active sites
12	Protein sequence databases - Protein domain databases; protein classification
13	Protein sequence databases - Databases of individual protein families
14	Structure Databases
15	Structure Databases - Small molecules
16	Structure Databases - Carbohydrates
17	Structure Databases - Nucleic acid structure
18	Structure Databases - Protein structure
19	Genomics Databases (non-vertebrate)
20	Genomics Databases (non-vertebrate) - Genome annotation terms, ontologies and nomenclature
21	Genomics Databases (non-vertebrate) - Taxonomy and identification
22	Genomics Databases (non-vertebrate) - General genomics databases
23	Genomics Databases (non-vertebrate) - Viral genome databases
24	Genomics Databases (non-vertebrate) - Prokaryotic genome databases
25	Genomics Databases (non-vertebrate) - Unicellular eukaryotes genome databases
26	Genomics Databases (non-vertebrate) - Fungal genome databases
27	Genomics Databases (non-vertebrate) - Invertebrate genome databases
28	Metabolic and Signaling Pathways
29	Metabolic and Signaling Pathways - Enzymes and enzyme nomenclature
30	Metabolic and Signaling Pathways - Metabolic pathways
31	Metabolic and Signaling Pathways - Protein-protein interactions
32	Metabolic and Signaling Pathways - Signalling pathways
33	Human and other Vertebrate Genomes

34	Human and other Vertebrate Genomes - Model organisms, comparative genomics
35	Human and other Vertebrate Genomes - Human genome databases, maps and viewers
36	Human and other Vertebrate Genomes - Human ORFs
37	Human Genes and Diseases
38	Human Genes and Diseases - General human genetics databases
39	Human Genes and Diseases - General polymorphism databases
40	Human Genes and Diseases - Cancer gene databases
41	Human Genes and Diseases - Gene-, system- or disease-specific databases
42	Microarray Data and other Gene Expression Databases
43	Proteomics Resources
44	Other Molecular Biology Databases
45	Other Molecular Biology Databases - Drugs and drug design
46	Other Molecular Biology Databases - Molecular probes and primers
47	Organelle databases
48	Organelle databases - Mitochondrial genes and proteins
49	Plant databases
50	Plant databases - General plant databases
51	Plant databases - Arabidopsis thaliana
52	Plant databases - Rice
53	Plant databases - Other plants
54	Immunological databases
55	Others (Free description)

8 利用許諾決定

最後に、アーカイブファイル及び簡易検索サイトで公開するデータ（以下、「データベース」と総括）に対しての利用許諾を決定し、NBDC が指定した Web フォームに入力します。NBDC では利用許諾テンプレートを用意していますので、特に法律の知識がなくても利用許諾を簡単に入力することができます。

なお、利用許諾テンプレートを使用しない、または利用許諾テンプレートに記載されていない利用許諾を追加する場合には NBDC までご相談ください(9.お問い合わせ先 参照)。

8.1 利用許諾テンプレート

利用許諾テンプレートは、クリエイティブ・コモンズ・パブリック・ライセンス「表示-継承」に準拠した標準利用許諾を採用した利用許諾のテンプレートです。標準利用許諾は、データベースの利用者に許可されていることと、利用者が守るべき条件の組み合わせです。利用許諾テンプレートは、

<http://dbarchive.biosciencedbc.jp/tmpl-std-license.html> から参照することができます。

■ 利用許諾テンプレート 第4版 (2015年11月17日)

寄託データベースの利用許諾は、通常、NBDC標準利用許諾（「クリエイティブ・コモンズ 表示-継承4.0 国際」を推奨）を使用します。利用許諾を作成する際の参考として、以下に利用許諾テンプレートを示します。

■ 本データベースの利用許諾

本データベースは、以下で定める利用許諾に基づきご利用いただくことができます。本利用許諾は、本データベース利用における許諾内容、及び利用者が従うべき条件を定めています。



本データベースの利用許諾は、クリエイティブ・コモンズ 表示-継承4.0 国際の定める利用許諾です。

本データベースのクレジットは、「[データベース名] © [作成者名(作成機関名)] licensed under CC表示-継承4.0 国際」ですので、利用にあたり必ず表示してください。

クリエイティブ・コモンズ 表示-継承4.0 国際の概要は [こちら](#)です。具体的な許諾条項は [こちら](#)をご覧ください。

本データベースにおいて、以下の条件に従う限り許諾されている事項：

1. 本データベースの全部または一部に自由にアクセスし、データを取得することができます。
2. 本データベースの全部または一部のデータを自由に再配布することができます。
3. 本データベースの全部または一部のデータを利用した、データベースなどの翻案物を自由に作成し、配布することができます。

本利用許諾に基づいて利用する際に従うべき条件：

1. 本データベースの全部または一部、あるいは翻案物の配布に際しては、本データベースの作成者のクレジットを表示しなければなりません。
2. 本データベースの全部または一部のデータを利用して作成された翻案物は、CC表示-継承4.0（もしくは、それ以降のバージョン）、またはCC表示-継承互換ライセンス（リストは[こちら](#)）の下で配布されなければなりません。
3. 本利用許諾で許諾されていない事項については、以下のデータベース作成者に連絡をとり、利用許諾を求める必要があります。

データベース作成者連絡先：

(住所)

(機関名・部署名・担当者名等)

E-mail:(メールアドレス)

図 8-1 利用許諾テンプレート (日本語版)

License to Use This Database

Last updated : 2015/11/17

You may use this database in compliance with the terms and conditions of the license described below. The license specifies the license terms regarding the use of this database and the requirements you must follow in using this database.



The license for this database is specified in the [Creative Commons Attribution-Share Alike 4.0 International](#).

If you use data from this database, please be sure attribute this database as follows: "[Database name] © [Creator] ([Organization]) licensed under CC Attribution-Share Alike 4.0 International".

The summary of the Creative Commons Attribution-Share Alike 4.0 International is found [here](#).

With regard to this database, you are licensed to:

1. freely access part or whole of this database, and acquire data;
2. freely redistribute part or whole of the data from this database; and
3. freely create and distribute database and other adapted materials based on part or whole of the data from this database,

under the license, as long as you comply with the following conditions:

1. You must attribute this database in the manner specified by the author or licensor when distributing part or whole of this database or any adapted material.
2. You must distribute any adapted material based on part or whole of the data from this database under CC Attribution-Share Alike 4.0 (or later), or CC Attribution-Share Alike Compatible License (the list is [here](#)).
3. You need to contact the Licensor shown below to request a license for use of this database or any part thereof not licensed under the license.

[contact person]

[organization]

[contact address]

TEL: [telephone]

E-mail: [email]

About Providing Links to This Database

You can freely provide links to all contents in this database. But, contents might be changed without notice.

図 8-2 利用許諾テンプレート(英語版)

8.2 標準利用許諾について

NBDC の定める標準利用許諾は、クリエイティブ・コモンズ・ジャパンの提供するクリ

クリエイティブ・コモンズ・パブリック・ライセンス「表示・継承」に準拠しています。標準利用許諾の中で使用される重要な用語の定義については、NBDC が作成した「データベース利用許諾関連用語の定義」

(http://dbarchive.biosciencedbc.jp/files/nbdc_license_glossary.pdf) (以下、「定義」) を参照してください。

8.2.1 クリエイティブ・コモンズ・ジャパンについて

クリエイティブ・コモンズ・ジャパンは、法律や技術に関する専門的な知識がなくても、簡単な要素条件の組み合わせを選択するだけで、誰でも著作物を、自分の希望する条件で、インターネットを通じて公開することができるライセンスシステムを提供しています。クリエイティブ・コモンズ・ジャパンに関する詳細な情報は、<http://www.creativecommons.jp/> を参照してください。

8.2.2 利用者に許可されていること

標準利用許諾においては、データベースの利用者は以下の行為が許可されます。

- データベースの複製及び再配布
- データベースの改変（「定義」を参照）
- データベースをダウンロードすることなく検索または閲覧する行為（著作物をダウンロードせずに単に閲覧するだけの行為は、著作権法の規制がされていません）。

8.2.3 利用者が守るべき条件

標準利用許諾においては、データベースの利用者は以下の条件を守らねばなりません。

- データベースを再配布する場合には、データベースの権利者（団体）の氏名、データベース名、指定 URL など、データベースに関する情報を表示すること。（「表示」）
- データベースを改変する場合には、オリジナルのデータベースの権利者（団体）が表示を求める情報を表示する。
- データベースを改変する場合には、オリジナルのデータベースの利用許諾と同一の利用許諾（つまり、今回の場合にはクリエイティブ・コモンズ・ライセンスの「表示・継承」）を、改変したデータベースにも適用すること。（「継承」）

8.2.4 データベースの営利利用及び改変について

クリエイティブ・コモンズ・ジャパンには、上記の「表示」・「継承」の他にも「非営利」・「改変禁止」という2つの要素条件が用意されています。しかしながら、NBDC では標準利用許諾を定めるにあたり、「非営利」・「改変禁止」を採用せず、データベースの営利利用

と改変を可能にしました。NBDC では、これにより DB 寄託者の努力が、科学の発展に大きなインパクトを与え得ると考えています。その理由は以下の通りです。

● データベースの営利利用について

クリエイティブ・コモンズ・パブリック・ライセンス日本版では、営利活動については自由に利用することを認めない「非営利」限定の条件を採用することが可能です。この「非営利」の選択肢は、主として、私財を投じて商業的なコンテンツを創作した創作者に対して、その投下資本を回収する機会を保障する目的で設けられました。

しかしながら、公的資金の活用によって得られた科学データの成果については、研究主体が投下資本の回収を行うことは必ずしも求められていません。むしろ、その研究成果を、営利企業を含めたすべての研究主体が平等に研究開発に利用できるようにすることにより、DB 寄託者のデータベースは、DB 寄託者のみでは実現できない大きなインパクトを世の中に与える可能性を持つことになります。

また、一般に公開されたデータベースに、仮に企業が他の機能を付加して有償で販売したとしても、無償で公開されている元のデータベースの所在についての情報源を明示したうえで行うのであれば、自由な機能付加競争であると見ることができ、利用者により広い選択肢を与えていることになって、バイオ情報産業が活性化すると考えることができます。同様のことは、オープン・ソース・ソフトウェアの分野でも広く行われていて、多くの営利企業が、公開したソフトウェアのソースコードの営利目的での利用を許容しています。バイオ情報産業においてもこのような活性化を促すために、データベースの営利利用を許すことが望ましいと考えられます。

● データベースの改変について

クリエイティブ・コモンズ・パブリック・ライセンス日本版は、主に「文学、絵画その他の芸術」の分野で用いられることを念頭において制度が作られています。これらの「文学、絵画その他の芸術」の分野は感性が重要な分野であり、それぞれの作品が誕生時に、作品として完成していると考えられることもできるため、「リミックス」(改変)が「改善」でないこともありえます。そのため、改変を制限する「改変禁止」の選択肢が設けられています。

一方 DB 寄託者のデータやデータベースは「常に進歩する科学」の一部です。科学の発展は、データやデータベースに「付加修正などを加えて再利用していく」ことによって支えられており、その意味で、改変はデータベースが科学に活かされる為の本質的な作業のひとつです。さらに、改変を許容することによって科学研究の継続的な発展に貢献すれば、貢献者としてクレジットされることにもつながります。以上のように、データベースを科学研究の進歩に生かすためには、改変を許容することが重要であると考えられます。

8.3 標準利用許諾を利用しない場合

もしも標準利用許諾が寄託者の希望に合わない場合には、標準利用許諾を使用せず、独自の利用許諾を作成することも可能です。その場合には NBDC にご相談ください（9.お問い合わせ先 参照）。

8.4 利用許諾の入力

利用許諾が決定したら、その内容を NBDC から指定された Web フォームに入力します。Web フォームには、予め利用許諾テンプレートが入力されていますので、多くの場合はテンプレートを編集するだけで済みます。

9 お問い合わせ先

生命科学系データベースアーカイブや本ガイドラインについてのご質問、アーカイブ化にあたってのご相談等は、以下の連絡先までお問い合わせください。

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
データベース寄託担当

〒102-0081 東京都千代田区四番町5-3 サイエンスプラザ

TEL 03-5214-8491, FAX 03-5214-8470

E-mail: dbarchive@biosciencedbc.jp

URL: <http://dbarchive.biosciencedbc.jp/>