



平成27年1月8日、15日
募集説明会

ライフサイエンスデータベース統合推進事業 「統合化推進プログラム」

プログラムの概要及び 募集・選考・プログラム運営の方針説明

研究総括

長洲 毅志

(エーザイ株式会社 プロダクトクリエーション本部 ポートフォリオ戦略・推進部 顧問)



ライフサイエンスにおける情報爆発

- データ爆発
 - 超高速ゲノム配列決定装置 100億文字/台/日
 - コンピューターの性能の伸び(ムーアの法則)を凌駕
 - 1万を超える生物のゲノムプロジェクト進行中
 - 画像、動画データも急増
- 知識爆発
 - 論文数 2,000万件
 - フルペーパーも利用可
- データベース爆発
 - 世界 1万
 - 日本 1千
 - 解析ツール 2千
 - 内容も多様化

ライフサイエンスはpeta(10^{15} =
千兆)オーダーのデータ量の
時代に突入

ライフ分野におけるデータの共有・統合の重要性

- データベースを指向した種々の大規模プロジェクト
 - 多額の研究費がデータに化けている
- 仮説駆動型からデータ駆動型(データ中心科学)へ
 - 法人化、バイドール法 → データの囲い込み
- データや知識の分断、断片化
- イノベーションには共有・統合化必要
 - ブラウズ(検索)や一部のデータ公開では不十分
 - データ生産者以外の不特定多数のイノベーターの存在
 - 再利用、転用、マッシュアップが自由にできることが不可欠



ライフイノベーションのために

- データ共有、統合によるオープンイノベーションが不可欠
 - ブラウズ(検索だけでは)不十分
 - 狭い意味の公開だけではだめ
 - データ生産者以外の不特定多数のイノベーターの存在
 - 再利用、転用、マッシュアップが重要
 - そのためにはコピー、再配布が自由に
 - 巨人の肩の上に立つという精神で

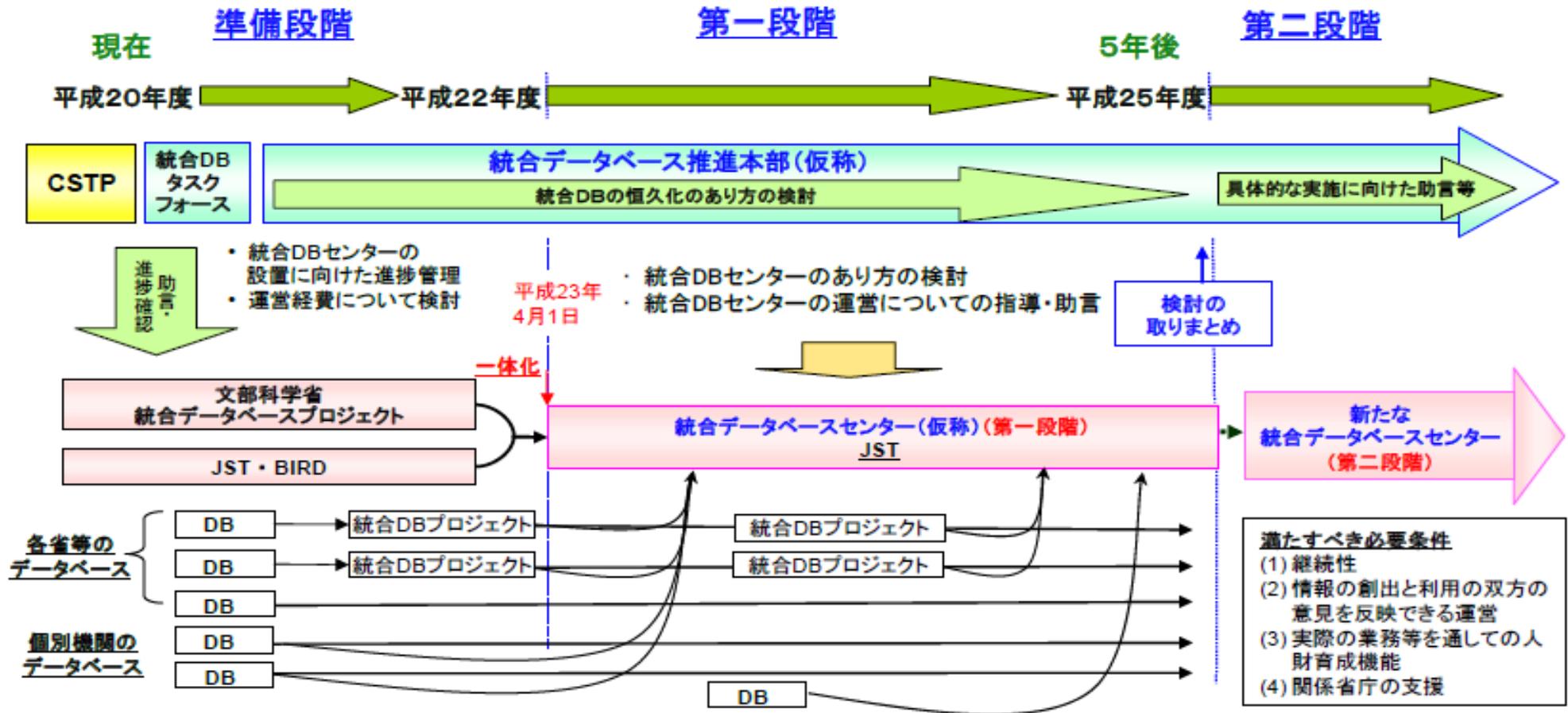


我が国におけるライフサイエンスDBの問題点

- DBが散在していて**所在情報**や**利用法**が分からない
 - 我が国で1,000、世界で10,000ものDB
 - 似たようなものがいくつもありどれを使ってよいか分からない
- DBやDBのエントリに信頼性の高い**注釈**がついていない
 - DB構築、維持を行える人材不足、DB構築への評価の低さ
- 大型プロジェクトの成果公開、共有が不十分
 - 我が国のライフ関係予算は年間3,000億円以上
 - 公開されているものも**プロジェクトが終了**すると更新ストップ
- **ばらばら**に構築・管理されていて検索・解析・応用が困難
 - 現在の統合化は分子レベルで行われていて医療などへの応用困難
 - 日本語化されていないので研究動向や分野の状況の把握困難
- 我が国にはDB戦略、中核センターがない



統合データベース整備のロードマップ



総合科学技術会議統合DBタスクフォース報告書(H21.5.21)より



文科省統合データベースプロジェクトポータルサイト

Guest Account | アカウント | ログイン | English

文部科学省委託研究開発事業

統合データベースプロジェクト

ホーム DBカタログ 横断検索 ツール ダウンロード About us

統合ホームページへようこそ

バイオサイエンスデータベースセンター

National Bioscience Database Center

- 散在するデータベースを、まとめて、使い易く -

科学技術振興機構
JST Japan Science and Technology Agency

文字サイズ変更 大 中 小

English [サイトマップ](#)

ホーム NBDCについて 研究開発プログラム 公募情報 採用情報 広報 お問い合わせ リンク

NBDCは、日本の生命科学研究を推進するために、データベースをつなげて使い易くします。そのためにNBDCや協力機関は、以下のようなサービスやウェブサイトを作成・提供しています。

生命科学全体のデータベース統合

[Integbioデータベースカタログ](#)

データベース横断検索

[生命科学系データベースアーカイブ](#)

分野ごとのデータベース統合

ヒトと医・薬

[NBDCヒトデータベース](#)

[ヒトゲノムバリエーションデータベース](#)

[ヒト疾患脳画像データベース](#)

[KEGG MEDICUS: 疾患・医薬品統合リソース](#)

生命を支える分子

[DDBJ: 日本DNAデータバンク](#)

日本語や動画でわかりやすく

[新着論文レビュー / 領域融合レビュー](#)

[統合TV](#)

論文をもっと読みやすく、書きやすく

[Allie / inMeXes / TogoDoc](#)

大量の配列データを扱いやすく

[DBCLS SRA / 鎖鋸 \(β\)](#)

[RefEx / 統合遺伝子検索 GGRNA](#)

さまざまな統合コンテンツ

[生物アイコン](#)

[生命科学系主要プロジェクト一覧](#)

NBDCパンフレット

(PDF: 2.65MB / 2013/04/08更新)

新着情報 [twitter](#) [RSS](#)

2013/12/25

【メンテナンス】2013年12月26日(木) 10:00~12月27日(金) 10:00の間、生命科学データベース横断検索の一部のデータベースが検索できなくなります。

2013/12/25

平成26年度ライフサイエンスデータベースセンター「統合化推進プログラム」の募集説明会を追加しました。

2013/12/25

【メンテナンス】2014年1月17日(金) 18:00~1月20日(月) 12:

6



主なサービスのデータベース収録数とアクセス数

生命科学系データベースカタログ

Integbio データベースカタログ

1325 件

1,381DB

カテゴリー別ツリ表示

サムネイル

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

検索結果

外部リンク

462DB

生命科学系データベースアーカイブ

LSSB Archive

生命科学系データベースアーカイブ

アーカイブの概要説明

アーカイブデータベース一覧

アーカイブデータベース一覧 (ヘルプ)	サムネイル、名称	運用場所、代表者	カテゴリー、生物種	要約、利用許諾条件など	収録数
Q-TARO	Q-TARO	農業生物資源研究所	動物	...	92DB
AT Atlas	AT Atlas
TP Atlas	TP Atlas

約400万ページビュー／年

(平成26年12月現在)



各省のデータベース統合

バイオサイエンスデータベースセンター
(文部科学省)

医薬基盤研究所
(厚生労働省)

農業生物資源研究所
(農林水産省)

バイオメディシナル情報研究センター
(経済産業省)

RDFによる
すべてのDBの統合

連邦型統合DB

③DB統合化基盤技術の開発

DB統合化基盤技術の開発

目的別統合DB

微生物DB

植物DB

ヒトDB

データバンク的DB

ゲノム

タンパク質立体構造

メタボローム

糖鎖

DBをまとめることで使
いやすく

統合化推進プログラム
(10課題)

④国内バイオ関連DB統合の促進

個々のDBを
使いやすく

カタログ

横断検索

アーカイブ

標準化DB3
標準化DB4b

②ポータルサイトの構築・運用

データベース

DB1a

DB1b

DB2

DB3

DB4a

DB4b

プロジェクト

P1

P2

P3

P4

測定装置

顕微鏡

MRI

chip

NGS

質量分析

NMR

統合化のロードマップ

第一期中（平成25年度まで）に実施予定

(1) ポータル構築連携

「生命科学系データベースカタログ」への記載、リンクの実施

新センターのポータルサイトに、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

「横断検索サーバー」の設置、一括横断検索の実現

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

「統一フォーマット」でのダウンロードの実現

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。

(4) データベース構築連携

「データベースの再構築」による統合の実現

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。

第二期以降を想定

本プログラムの目的(全体の考え方)

(募集要項 8ページ)

本プログラムは、国内外に散在しているライフサイエンス分野のデータやデータベースの共有を強力に促進し、公共財として誰でもが自由に活用できるようにするとともに、生物種やプロジェクトを超えて幅広い統合化を実現することにより、データがより多くの分野の研究者、開発者、技術者に簡便に利活用できるようにして、データの価値を最大化することを目指すものです。すなわち、データの共有、統合を通して、我が国のライフサイエンス研究の成果をあまねく行き渡るようにするとともに、それを十分に活用できる環境を構築することにより、ライフ分野におけるイノベーションを促すことが本プログラムの目的です。

対象となる研究開発提案

(募集要項 8ページ)

- 個別研究分野をまたいで、高い汎用性(分野を超えて多くの研究者、開発者、技術者が利用可能で自分の仕事に役立てられるものであること)があるデータをできるだけ**網羅**しているもの
- フォーマットやオントロジーが**標準化**されていて、統一的なインタフェースで誰でも簡単に利用できる統合データベースの構築を目指すもの
- 我が国の中核、拠点となるデータベースであることはもちろんのこと、世界的にみても十分な**競争力**、発信力を備えた日本を代表するデータベースを目指すもの
- プロジェクトや機関を超えて、また、個別の生物種や個別の生命現象の枠をできるだけ超えて、データを収録・統合する**横断的**データベースであること

なお、データベースに入れるべきデータの産生を目的とした活動は、本プログラムの対象ではありません。

研究開発内容(1)

(募集要項 9ページ)

- 本プログラムの研究開発対象となるデータベースは、本プログラムの過去3年間で開発されてきた統合データベースの質的量的にさらなる充実を図るものや、それらの統合データベースと**補完的**な役割をもつデータベースであることが求められます。

例えば、

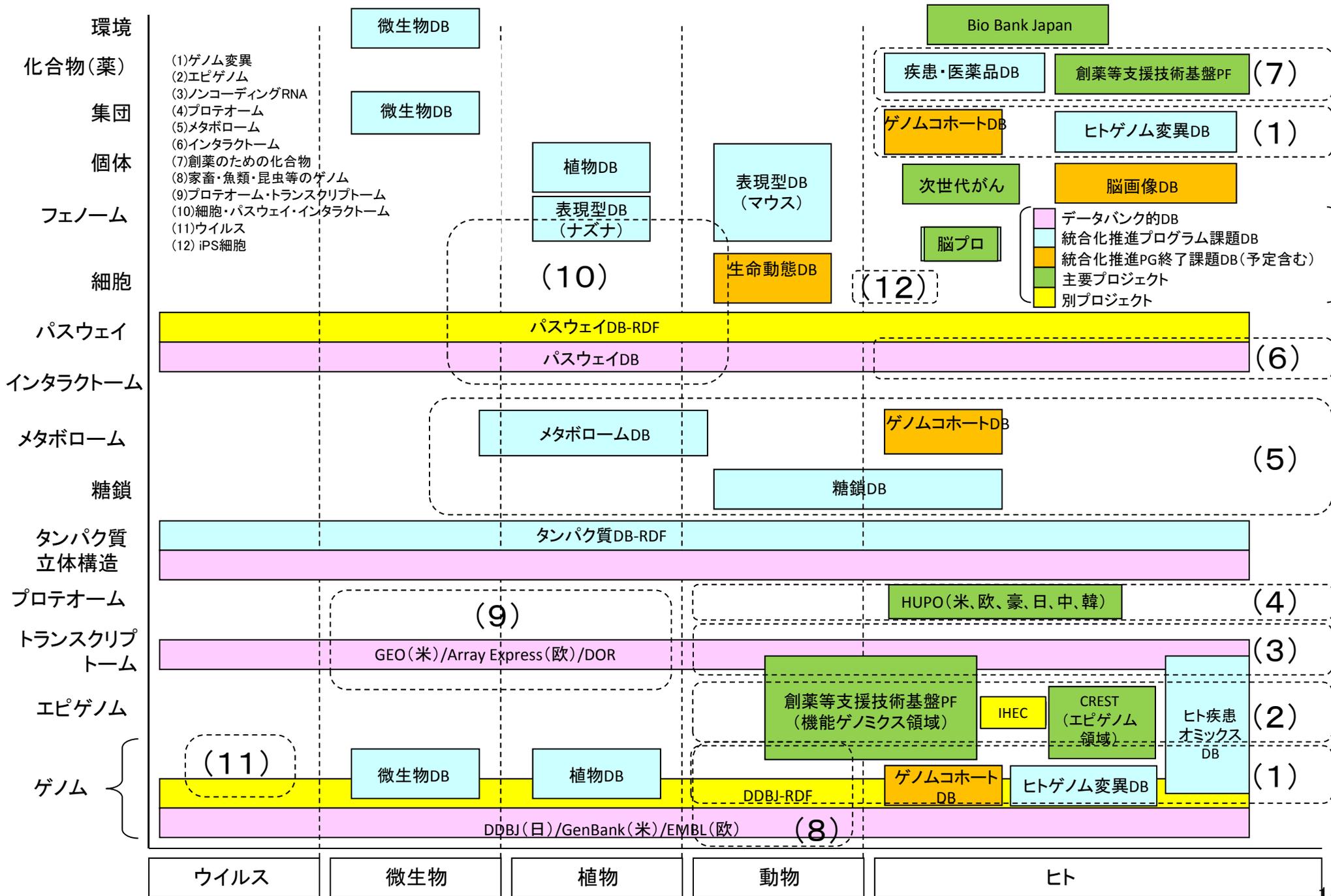
ヒトのエピゲノム、ノンコーディングRNA、
プロテオーム、インタラクトーム、

メタボローム、および、創薬のための化合物等

植物では、プロテオーム、微生物ではメタボローム等のデータベース

さらに、(食料生産としての観点から)家畜や魚類、昆虫等のゲノム





研究開発内容(2)

(募集要項 10ページ)

- データの汎用性の観点から
ゲノム配列やタンパク質立体構造等の分子レベルデータの統合化、
加えて、基礎分野だけでなく、医学や農学等への応用、イノベーションや
産業につながるデータの統合化も歓迎
- データの網羅性の観点から
プロジェクトや機関を超えたデータの収録をすること、さらに、
データバンクのような登録機能を実現することも本プログラムの範疇
- データの共有の観点から
学界だけでなく、産業界からも無償で自由に利用出来ること
データ共有、公開のルールが明確であること
出来るだけ早い時期の公開、共有をすること
- その他
出来るだけ、自動化、省力化する等して、持続可能性の高いものであること

研究開発提案が満たすべき条件(1)

(募集要項 10ページ)

＜データの公開＞

- 本プログラムで構築されたデータベースは基本的にすべて無償で公開、ダウンロード可能であること
- NBDCのアーカイブサイト(<http://dbarchive.biosciencedbc.jp>)から無償で提供可能であること
- 原則として、産業界からも学界からも、制約なく無償で利用可能であること。より具体的には、クリエイティブコモンズのCC BY-SA で提供可能であること
- ヒトに関するデータ等の機微情報に関しても、制限付きアクセスで利用可能であること
- 本プログラムの研究開発課題に関連して研究代表者グループ及び共同研究グループが保有している全てのデータ(過去に取得したデータも含まれる)について、本プログラムで構築されたデータベースに格納し、公開されること

研究開発提案が満たすべき条件(2)

(募集要項 11ページ)

＜データの補完性＞

- 現在の統合化推進プログラムで構築されているデータベースやその他の国内外のデータベースとの補完性、接続性があり、それにより、より高い統合化が実現できるものであること

＜データの汎用性＞

- 本プログラムで研究開発するデータベースに関して、個々の研究分野やプロジェクトの枠を超えて、十分な数の利用者(例えば、一つの目安として数万程度)が見込まれること

＜データベースの国際競争力＞

- 十分な国際競争力を備えていること(外国からも十分なアクセスが見込まれるものであること)。具体的には、外国にある類似のデータベースと遜色ないアクセス数が見込まれること

研究開発提案が満たすべき条件(3)

(募集要項 11ページ)

＜データの標準化＞

- データ統合対象領域におけるデータ収集のフォーマットやオントロジーの標準化に向けて実効的な計画を有していること
- データベースの構築、統合化に関して、NBDCから提供されるガイドライン、フォーマット、オントロジー(例:NBDCヒトデータ共有ガイドライン、NBDCヒトデータ取扱いセキュリティガイドライン)がある場合は、それに従うこと

＜データの管理体制＞

- データベースの開発に際しては、データの管理に十分な体制、システムが構築されている(構築される予定である)こと。特にヒトに由来するデータ等の機微情報の扱いには特別の配慮をすること

研究開発提案が満たすべき条件(4)

(募集要項 11ページ)

＜ NBDC への協力＞

- NBDCで構築あるいは提供される他のデータベースとの連携に協力することに加え、NBDCが目指す統合データベースの方向性に従い、研究開発を進めること
- データベースのアクセス情報およびその活用情報を収集し、NBDCに提供すること
- NBDCで取り組むIntegbioデータベースカタログ、生命科学データベース横断検索、生命科学系データベースアーカイブ、さらにヒトデータに関しては、NBDCヒトデータベースへのデータの提供に協力すること
- NBDCが実施する公開シンポジウム、利用者アンケート、学会等の展示会、講習会、パンフレット作り等に協力すること