



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(統合データ解析トライアル)」

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

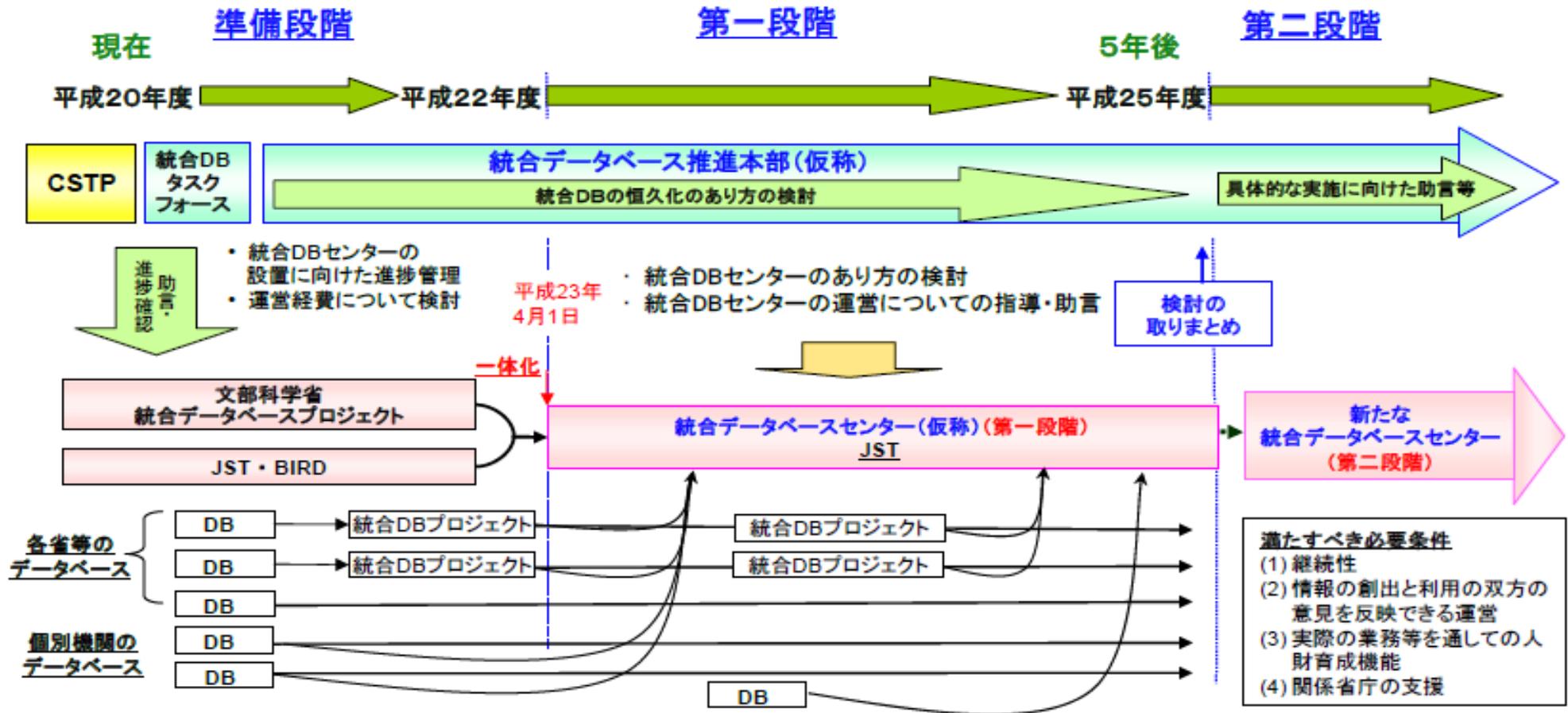
- データ共有、統合によるオープンイノベーションが不可欠
 - ブラウズ(検索だけでは)不十分
 - 狭い意味の公開だけではだめ
 - データ生産者以外の不特定多数のイノベーターの存在
 - 再利用、転用、マッシュアップが重要
 - そのためにはコピー、再配布が自由に
 - 巨人の肩の上に立つという精神で



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 -

JST 独立行政法人 科学技術振興機構
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

ポータル

Integbioデータベース
(旧)生命科学系データベース
生命科学系主要データベース
生物アイコン
ライフサイエンス
ライフサイエンス
WingPro (JSTの)
Webリソースポ

検索

生命科学データ
蛋白質核磁気共

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](#)

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](#)

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](#)

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:](#)

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

462DB

外部リンク

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

92DB

アーカイブの概要説明

新アーカイブ情報

アーカイブデータベース一覧

データベース	データベース運用機関	代表者	運用場所	生物種	利用許諾	
Q-TARO	ダウロード 福井県総合リサーチセンター	農業生物資源研究所	丸丸 博一	QTL	イヌ	立ち寄り系のため30日QTLデータベース
AT Atlas	ダウロード 福井県総合リサーチセンター	文部科学省 超ゲノムバンク研究プログラム	野原 英樹	その他の分子生物学データベース	植物学系	超ゲノムバンク研究プログラム関連された検索のデータベース
TP Atlas	ダウロード 福井県総合リサーチセンター	文部科学省 超ゲノムバンク研究プログラム	野原 英樹	代謝系	各種モデル	ネットワーク図をプラットフォームと超ゲノムバンクの研究成果

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)

バイオメディカ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

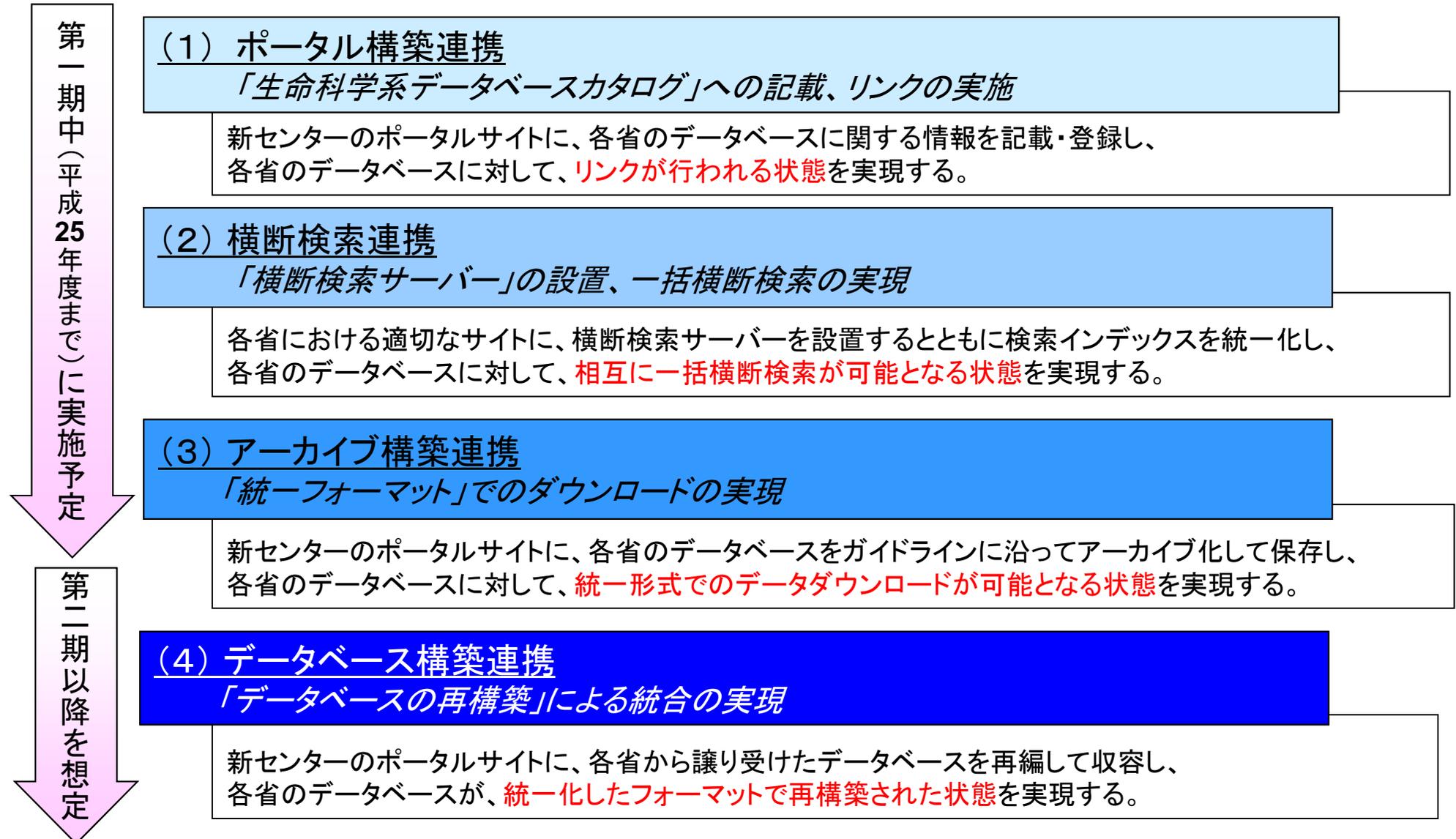
chip

NGS

質量分析

NMR

統合化のロードマップ



生命科学系データ合同ポータルサイト

文字サイズ

integbio.jp 生命科学系データベース統合のための合同ポータルサイト

ホーム 統合へのステップ 技術交流 イベント

4省の生命科学系データベースの統合を目指して

このサイトは、文部科学省、厚生労働省、農林水産省、経済産業省による、生命科学系データベース統合のための合同ポータルサイトです。

内閣府 総合科学技術会議ライフサイエンスプロジェクトチーム 統合データベース(推進)タスクフォースの示す方針を踏まえて、平成23年度より、4省の間で生命科学系データベースの統合が推進されています。その統合のためのステップとして、データベースのカatalog、横断検索、アーカイブ構築などの連携を4省合同で進めています(図1)。

(1) カタログ連携 : データベースカタログへの記載、リンクの実施
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
各省のデータベースもガイドラインに沿ってアーカイブ化して登録し、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) データベース再構築連携 : 「データベースの再構築」による統合の実現
各省から提供されたデータベースを再編して登録し、各省のデータベースが、**統一したフォーマットで再構築された状態**を実現する。

データベースを探す
一覧 /

データを探す

データを使う
一覧 /

図1. 4省データベース統合の段階的進展

以下は、各省で統合を進めている4つの機関それぞれのポータルサイトです。

文部科学省: **NBDC-バイオサイエンスデータベースセンター**

NBDC | バイオサイエンスデータベースセンター | NBDCが データベース整備・統合の様々な実務に推れるとともに、ファウンディング

本プログラムの目的(全体の考え方)

(募集要項 9ページ)

- 本トライアルは、統合化推進プログラムで統合された(または統合中の)データベースについて、データベース利用者視点での**新たな活用**方法を探ることを目的としています。
- そこで、本トライアルは、統合化推進プログラムのデータベースの**有用性を高める**ツールの開発を実施することを目標としました。具体的には、統合化推進プログラムで統合されたデータベースや他のデータベースを更に組み合わせるためのツール、統合化推進プログラムのデータベースに付加価値をつけるためのツール(新たなデータ抽出法、他のデータとのリンクの付与など)の開発などが想定されます。

主な統合化推進プログラムデータベース

(募集要項 10ページ)

- メタボローム関連データベース
- 疾患・医薬品・環境物質関連データベース
- 微生物関連データベース
- 植物関連データベース
- ヒトゲノム多様性関連データベース
- フェノーム関連データベース
- 蛋白質構造関連データベース
- 糖鎖関連データベース
- ゲノム疫学関連データベース
- バイオイメージ関連データベース

詳細は「研究開発成果データベースリンク」

<http://biosciencedbc.jp/tec-dev-prog/db-link> をご覧ください。

研究開発提案のタイプ

(募集要項 10ページ)

- 研究開発提案タイプ(A):

本トライアルの要件を満たす課題を自ら設定し、研究開発提案を行うものです。

- 研究開発提案タイプ(B):

NBDCが提示した具体的な課題に基づき、その課題を解決するためのツールについて研究開発提案を行うものです。

研究開発提案が満たすべき条件

(募集要項 11ページ)

- 統合化推進プログラムで統合された(または統合中の)データベースの**有用性を高めるツール**を開発するものであること。実験等による、データの算出を目的とした活動は対象外です。
- 作成したツールは、広く研究者に活用されるよう、**公開**すること。また、特別な事情を除き、ソースコードも公開することを期待します。
- まだ公開されていないデータを用いる場合、研究開発提案書においてそのデータ内容を明らかにするとともに、データの利用許諾を取得すること。また、得られた**解析結果は公開**すること。ただし、論文作成や特許出願のための猶予期間を設けたりすることは可能です。
- 本提案で使用するデータは、研究開発提案者自身で取得すること。

今年度の試み（願い）

Bioinformaticianが増えてほしい。
ほかの分野から入ってきていただきたい。

そのためには、

非専門分野に入るには2人チームはよいかもしれない。
→そのための資金措置

非専門領域の人にとって課題設定が難しいかもしれない
→研究開発提案タイプ(B)の新設

