JAIST Repository
https://dspace.jaist.ac.jp/

Title	日本のバイオインフォマティクス政策のマクロビュー：グラントデータに基づくインフラ型プロジェクトとアプリケーション型プロジェクトの利用関係の分析
Author(s)	柳澤，直宏；加納，信吾
Citation	年次学術大会講演要旨集，26：222-225
Issue Date	2011-10-15
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/10106
Rights	本著作物は研究・技術計画学会の許可のもとに掲載するものです。This material is posted here with permission of the Japan Society for Science Policy and Research Management.
Description	一般講演要旨

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

２Ｂ０９

日本のバイオインフォマティクス政策のマクロビュー ～グラントデータに基づくインフラ型プロジェクトとアプリケーション型プロジェクトの利用関係の分析～

○柳澤 直宏，加納 信吾（東京大学）

１.はじめに

1.1 サイエンス・インフラとしてのバイオインフォマティクス

バイオインフォマティクスは、生物学、計算科学、数学、統計学を用いて、生物学的配列データやゲノムの内容および編成を分析し、巨大分子の機能と構造を予測する、学際的な研究分野と定義されていた（Luscombe et al., 2001)。そして、これら各種研究をするために、ゲノムから得られる配列とその内容に関する情報、実験的に得られた遺伝子産物の構造、機能分析から得られる情報のデータベース化が行われてきた。

特に近年では、ゲノムから得られる情報の意味を理解するポストゲノム研究が盛んである。これは、ゲノム情報が細胞内要素の設計図にすぎず、このゲノム情報に基づいて多様なタンパク質が発現し、発現したタンパク質の化学反応によって代謝物質を変化させることで細胞の活動が生じるからである。このことから、タンパク情報の集合であるプロテオームのデータや、代謝物質情報の集合であるメタボロームのデータ等多種多様なデータがデータベースの形で蓄積されている。

日本における代表的なデータベースの一例として Kyoto Encyclopedia of Genes and Genomics（KEGG）（Kanehisa et al, 2006）が存在するが、このデータベースは文献を基に代謝パスウェイデータを構築公開しており、近年では薬物のデータベースを新たに追加し、その構造情報や標的分子の情報を知ることが可能である。そしてこの KEGG は、サイエンス・インフラとして、国内外の研究者に広く利用されている。このように、バイオインフォマティクスには、サイエンス・インフラとしての役割がある。

1.2 ライフサイエンス・データベースを統合する活動

しかしながら、日本には数百のデータベースが存在しているといわれているものの、その利用が十分といえないもの少なからず存在する。このような現状を踏まえて、文部科学省委託研究開発事業である統合データベースプロジェクトによって、生命科学系データベースのポータルサイトが整備され、各種データベースのチュートリアルが映像コンテンツ等によって提供されている（Kawano et al., 2011)。このような活動によって、データベースの存在が明らかになるとともに、その利用が容易になってきている。

1.3 インフラ型プロジェクトとアプリケーション型プロジェクトの定義

本研究においては、委託費、補助金、交付金等を含めたグラントを受けたライフサイエンスの研究プロジェクトのうち、データベースの構築、データの拡充を研究目的の一つとして掲げているものをインフラ型プロジェクトと定義する（バイオリソースのバンクに関するもの、統合データベースの構築に関するものを除く)。

また、バイオインフォマティクスの研究プロジェクトのうちデータベースの構築およびデータベースの拡充以外を研究目的とするものをアプリケーション型プロジェクトと定義する。尚、アプリケーション型プロジェクトには、データベースの構築と生命現象の分析をテーマとするインフラ型プロジェクトとアプリケーション型プロジェクトのハイブリット型を含むものとする。

２．研究目的

本研究の目的は、インフラ型プロジェクトとアプリケーション型プロジェクトの利用関係の分析することである。この分析は、日本におけるバイオインフォマティクス研究の評価、日本におけるバイオインフォマティクス政策の評価につながるものであると考えている。しかしながら、インフラ型プロジェクトとアプリケーション型プロジェクトとの間には直接的な利用関係がある訳ではないため、その評価は困難である。本研究においては、インフラ型プロジェクトの成果物であると共に、アプリケーション型プロジェクトの基盤となるデータベースを、インフラ型プロジェクトとアプリケーション型プロジェクトの結節点として捉え、インフラ型プロジェクトとアプリケーション型プロジェクトの利用関係の分析する試みを行った。

尚、統合データベースプロジェクトに代表されるように、国内に存在（散在）している数百のデータベースを統合する活動が活発化しているが、グラントプロジェクトとデータベースとの対応関係が明らかになっているものは国内の主要プロジェクトに限られている。また、図１に示すように、グラントプロジェクトとデータベースの対応関係には多様なパターンが存在しているため、データベースの構築にあたってどの程度の資金が投入されてきたかをまとめた情報は知られている限り存在しない。これは、グラントの種類が多いこと、実施省庁が複数存在することにその原因の一端があると考えられるが、このような複雑な状況において、本研究は、現時点で得られるデータを最大限用いてどの程度の検討が可能であるかを示すことも本研究の目的とする。

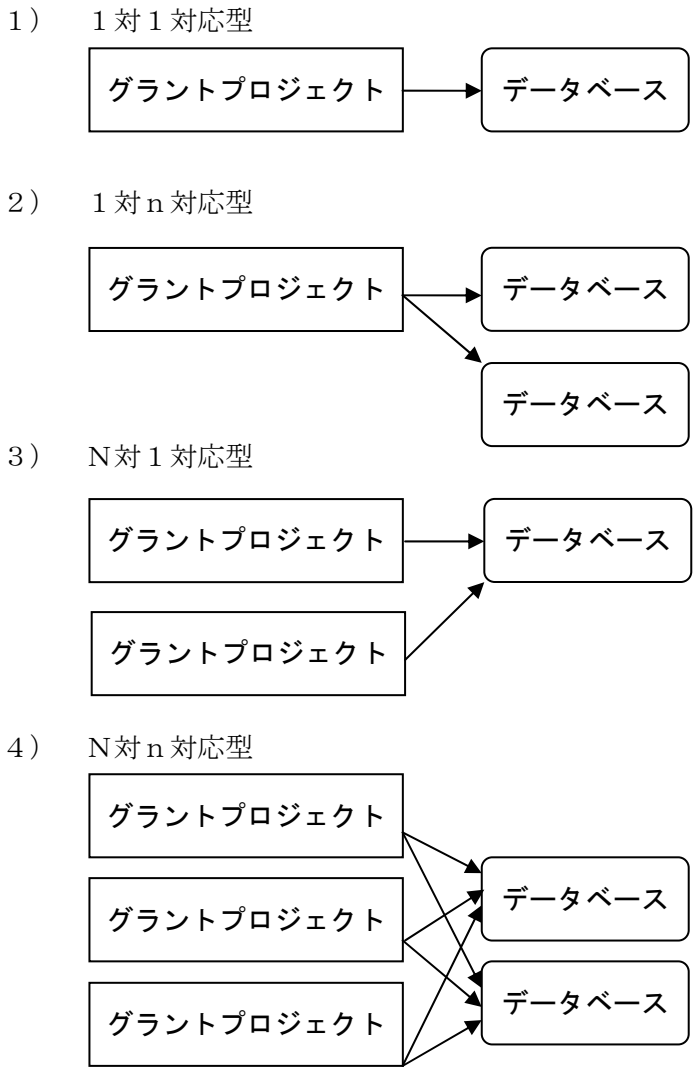

図１：類型化されたグラントプロジェクトとデータベースの対応関係

３．分析方法

3.1 インフラ型プロジェクトとデータベースとの対応付け

上記 1.3 で定義したインフラ型プロジェクトのうち、少なくとも一部の期間の配分額が確認できるもの（農林水産省の助成プロジェクト：１１件、厚生労働省の助成プロジェクト：４件、経済産業省の助成プロジェクト（NEDO、JBiC によるものを含む）：約２０件、文部科学省の助成プロジェクト（科研費、２１世紀 COE を含む）：約５０件）と、生命科学系データベースカタログにエントリされている約９００のデータベースのうち、国内の組織によって運営されているもの（組織分類されていないものを除く）との対応付けを行った。

3.2 データベースとアプリケーション型プロジェクトの対応付け

生命科学系データベースカタログにエントリされている国内外のデータベースとアプリケーション型プロジェクトとの関連性を分析した。尚、本分析においては、テキストマイニングの手法を利用している。具体的には、生命科学系データベースカタログに掲載されている国内外のデータベース約９００の紹介文とアプリケーション型プロジェクト約７００件の要旨との比較をすることによって、各データベースと各アプリケーションプロジェクトとの間にどの程度の関連性を網羅的に判定し、アプリケーション型プロジェクトとどの程度関連性があるかを示したスコア（関連度）を算出した。

3.3 インフラ型プロジェクトとアプリケーション型プロジェクトの利用関係の分析

上記 3.1、3.2 の結果を踏まえて、インフラ型プロジェクトと対応付けられた各データベースに配分されたと推定さる額と、各データベースとアプリケーション型プロジェクトとの関連度に基づいて、データベースを結節点としたインフラ型プロジェクトとアプリケーション型プロジェクトの利用関係を分析した（図２）。

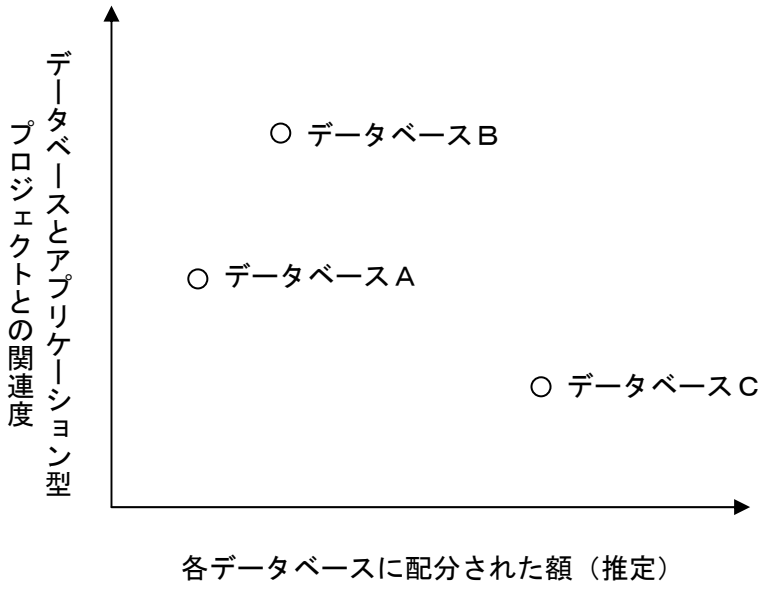

図２：インフラ型プロジェクトとアプリケーション型プロジェクトの利用関係

４．結果と考察

現在のところ、図２に示した、データベースを結節点としたインフラ型プロジェクトとアプリケーション型プロジェクトの利用関係について、年代別、グラントの種類別、プロジェクトの実施省庁別等の

切り口で分類し、各種検討を行っている。結果と考察の詳細については本発表にて紹介する。

５．おわりに

本研究は、基礎研究分野におけるグラントの評価を目的としたグラント・メトリクスに関するものとして、一定の成果があったと考えている。しかしながら本研究において、グラントプロジェクトとデータベースとの対応付けが確認されなかったものが少なからず存在していることが明らかになっている。すなわち、データベースの構築を研究プロジェクトの目的の一つとして掲げているものの、その成果物であるデータの所在が不明という状態である。このことから、本研究は、更なる調査の余地がある。従って、本研究は、バイオインフォマティクス分野におけるインフラ型プロジェクトとアプリケーション型プロジェクトの利用関係の一部しか反映できていないかもしれない。一方、現在得られるデータに基づいてどの程度の検討が可能であるかを示す目的は達成できたといえる。

尚、本研究を通じて、今後、グラントプロジェクトの成果報告書に、業績として学術論文を列挙するように、データベースの構築やデータの拡充等の情報を掲載する必要性を強く感じたため、ここに提言したい。

参考文献

Kanehisa M., Goto S., Hattori M., Aoki-Kinoshita K.F., Itoh M., Kawashima S., Katayama T., Araki M., and Hirakawa M.(2006) Nucleic Acids Res. **34**, D354-357

Kawano S, Ono H, Takagi T and Bono H.(2011) Brief Bioinform **12**,(Online Journal)

Luscombe N.M., Greenbaum D. and Gerstein M. (2001) Methods of Information in Medicine **40**, 346-358