

世界初のデータベース総合 インキュベーションセンター



豊田哲郎
(とよだ・てつろう)
部門長

ライフサイエンス分野では、研究サンプルを最初に丸ごと測定してデータベース化し、後からそのデータを詳しく分析していく「オミックス研究」というスタイルが主流となりました。このため、各研究ラボでは研究対象をモノとして扱うだけでなく、情報化されたデータベースとしても扱えることが不可欠です。さらに各ラボの主宰者は、それら多数のデータベースを仲立ちにして、他の研究者との国際的な連携関係を築いていかねばなりません。生命情報基盤研究部門では、このようなラボ環境を誰もが享受できるバーチャルラボセンター（理研サイエンス）をインターネット上に構築し、データベースの統合化までを繋ぐ国家プロジェクトを推進します。

無数のバーチャルラボをサイバースペース上に構築して研究者にレンタル

近年のライフサイエンス研究は、大規模データベースを核として遠隔地や海外の研究者とグローバルな連携関係を築きながら研究を進めていく点に特徴があります。これは、現実空間に存在するラボでの活動というよりも、空間を超越したバーチャルなラボ内に遠隔地から大勢が参画して、そこでデータや意見交換しながら共同研究しているイメージに近いといえます。実はこのバーチャルな関係こそが、研究活動体の本当の実態を表しており、“虚実の逆転現象”が起きていると言われていています。

現在、ほとんどの研究者は電子メールを駆使してこのバーチャルな活動を行っているため、電子メールではできないことが彼らの活動の限界となっています。例えば、大型のデータベースを電子メールで扱うことはできません。このため、各研究者はメール以上のことをしようとして独自の情報システムを個別に立ち上げて、自らの研究コミュニティを支えているのが現状です。しかし、この様にシステムを個別に構築することは、開発コストだけでなく、運

用コストまで各研究者の負担として長期的に重く押し掛かるため、研究者達の足かせになっています。

そこで当部門では、強力なデータベース構築機能、機密情報が管理できる電子ラボノート機能や、メール機能など、国際的な連携研究に不可欠な機能を具備し、高いセキュリティで堅牢に守られたバーチャルラボを無数に構築する技術を開発し、「理研サイエンス (RIKEN SciNeS: RIKEN Life Science Networking System)」という名称で、バーチャルラボの試験レンタルを開始しました。この理研サイエンスは、仮想的なデータだけでなく現実的な物品にいたるまで、すべてに固有のIDを割り当てて管理し、虚実の世界がシームレスに統合化されたバーチャルラボを提供します。理研サイエンスによって、バーチャルラボのオーナーとなった研究者は、そこでのバイオインフォマティクスや連携研究を主導できる優位な立場に立つことが可能になります。

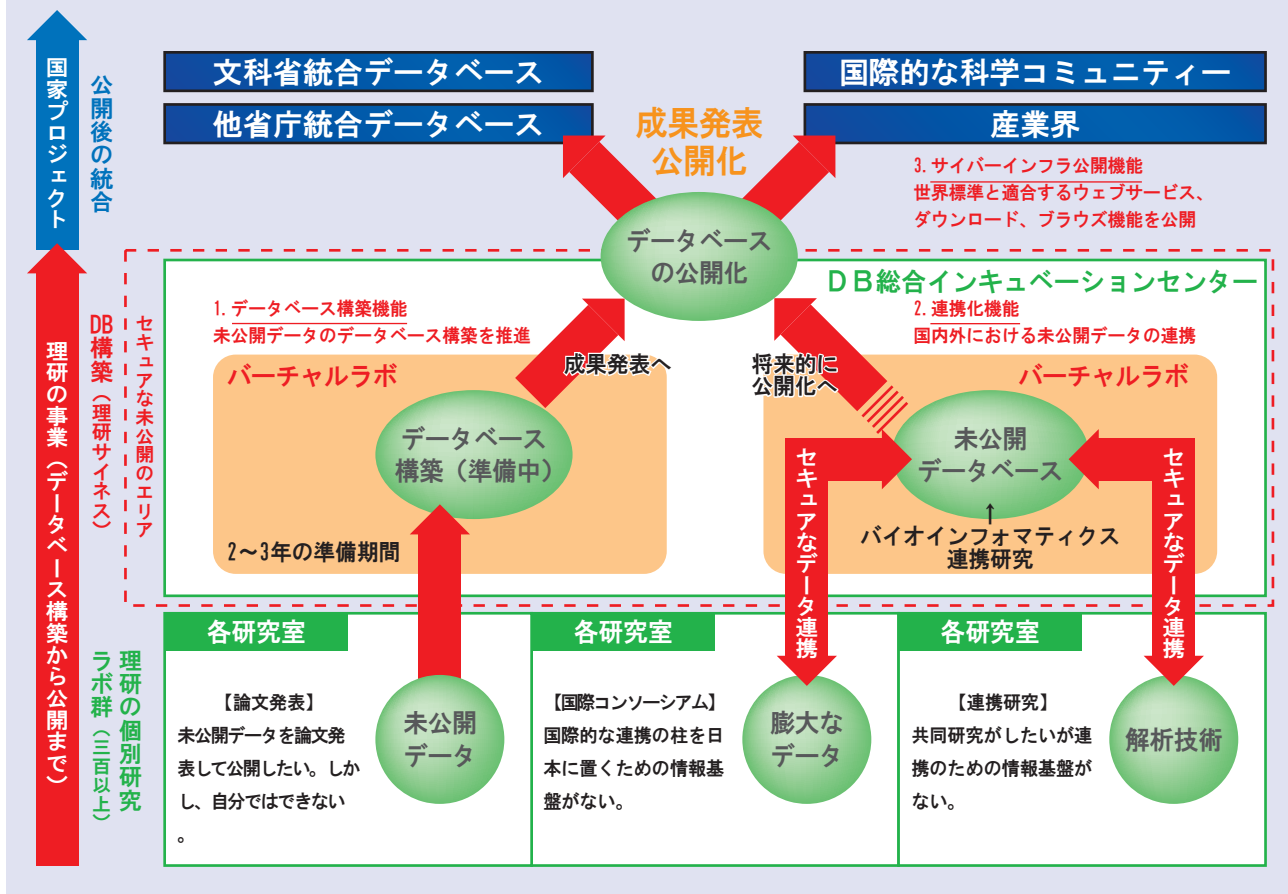
組織の壁を超えて連携力を強化する

これからの日本が総合力を発揮して国際的な競争に打ち勝っていくためには、組織の壁を超えて連携力を強化する必要があり、国内外の連携体制を機動的に構築できる我が国発の情報基盤が必要です。これまでの日本では、優れた情報基盤を国際的に提供することの意義が十分認識されておらず、日本の研究者が核となって主導権を発揮できる優れた連携システムを整備することが急務となっていました。理研サイネスは、国際標準規格であるセマンティックウェブ技術に準拠して開発された理研独自の技術であり、遠隔地に分散した

ライフサイエンス研究者たちを結びつける情報基盤として、今後はさらに、トランスレーショナルリサーチの分野における情報連携基盤としての利用に期待が高まります。現在、理研免疫・アレルギー科学総合研究センターと連携して、専門医と臨床医をつなぐネットワーク構築に理研サイネスの基盤を活用する計画です。また、理研バイオリソースセンターとの連携では、国際的なマウスリソース表現型データベースの統合化でも日本の連携拠点としての役割を理研サイネスが担っていく予定です。

【図1】 データベース総合インキュベーションセンター（理研サイネス）が提供する連携機能

生命情報基盤研究部門によるデータベース構築、連携、公開化の共通基盤（理研サイネス）データベース開発四段階（「構築」→「連携」→「公開」→「統合」）をカバーする総合基盤



データベース構築から公開化までのインキュベーション機能を理研サイネスが提供し、国際標準規格に準拠したインターフェースで公開することで、世界的なデータベース統合のためのサイバーインフラストラクチャーの形成にも貢献する。理研サイネスの試用版は <http://database.riken.jp/> から公開中。

データベースの構築から統合化までを円滑につなぐ

研究論文の成果発表では学術雑誌という専用の発表メディアが発達しているのに対し、データベースの成果発表では、個々の研究者が自らウェブサイトを立ち上げてサービスしなければならず、発表後も継続的にサービスを維持するための運用コストやデータベース統合化にかかるコストが大きな問題となっていました。理研サイネスはこの問題を根本解決する切り札として期待されます。

このため、理研サイネスは、個々の研究者がデータベースを発表するための「新しいタイプの学術メディア」となることが期待されており、この基盤から公開されたデータベースは統合化も容易であるため、現在、文部科学省が推進している統合データベースプロジェクトでも、理研サイネスを使って理研のデータベース公開化と統合化が推進されており、統合データベースセンターへのデータの受け渡しを円滑に行っています。

理研サイネスの特徴

- ①数万個以上の個別データベース構築活動を、大勢の研究者がインターネット経由で並行して実施できる。
- ②大規模なデータを介した業務フローを柔軟に設定でき、人的連携や自動処理を容易化できる。
- ③各活動群をセキュリティの高い状態で区切り、未公開の状態データベース構築ができる。
- ④構築したデータベースをその基盤から直接公開できる。
- ⑤公開後も、研究者がシステムの維持コストを負担することなく、その基盤でコンテンツを継続的に更新することができる。
- ⑥複数の世界標準形式に準拠したデータ配信が容易

【図2】 理想的な生命情報基盤に求められる機能 — (理研サイネスが有する機能)

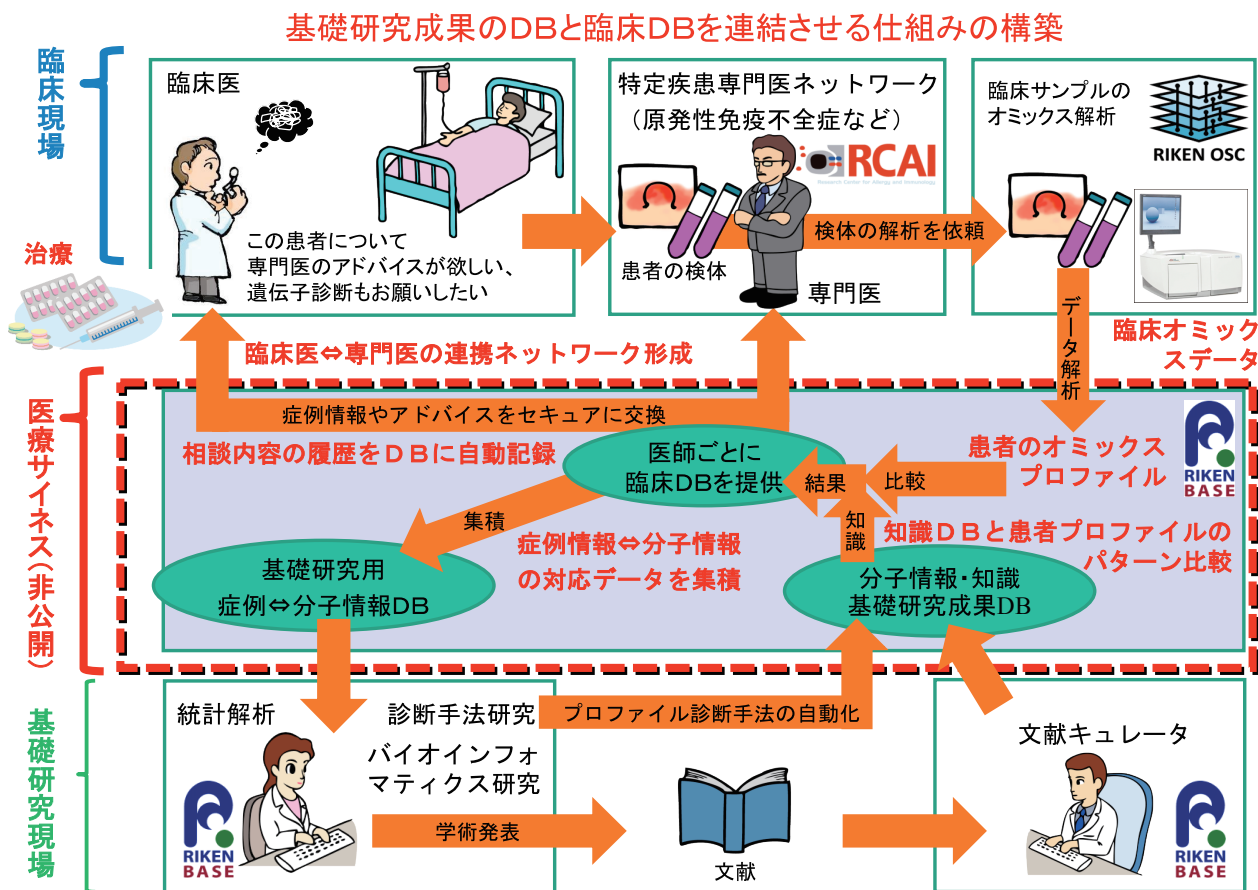
大規模データを介した連携研究のためのインキュベーション基盤

大規模データを介した共同研究のための理研サイネス “ライフサイエンスネットワーキングシステム”

連携化支援機能

- ◇コラボレーション・レビュー機能
- ◇メッセージ機能
- ◇電子ラボノート機能
- ◇コミュニティ機能
- ◇プロジェクト運営機能
 - オントロジープロジェクト
 - ・オントロジー構築機能
 - ・異種オントロジー間対応付け支援機能
 - データベースプロジェクト
 - ・データベース構築機能
 - ・データベース公開機能
 - ・文書タグ付けルール設定機能
 - ・自動文書タグ付け機能
 - ・Wiki機能
 - データベーステンプレートプロジェクト
 - ・データベースウィザード機能
- ◇リポジトリ機能
 - ・ライブデータベースリポジトリ機能
- ◇解析ツール拡張機能
 - ・データ自動解析機能
 - ・データ可視化機能
- ◇一括ダウンロード機能
 - ・データマート機能
- ◇データフロー制御機能
 - ・ラボラトリーオートメーション
- ◇自動更新機能
 - ・自動セマンティックウェブ変換機能
 - ・データベース自動統合更新機能
- ◇推論検索機能
 - ・推論検索エンジンGRASE (PosMed, etc.)
 - ・階層構造一括検索機能
 - ・塩基配列または化合物構造のようなキーワード以外の検索
- ◇安定した運用
 - ・スーパーコンピュータとの連携
 - ・さまざまなサービスの機能
- ◇災害対策
- ◇システムメンテナンスとサポートスタッフ

【図3】 医療サイネス（医療連携研究情報基盤）の提供計画（平成22年度～）



最近の成果紹介

- 【英文誌】 1. Yoshida Y., Makita Y., Heida N., Asano S., Matsushima A., Ishii M., Mochizuki Y., Masuya H., Wakana S., Kobayashi N., Toyoda T. PosMed (Positional Medline): prioritizing genes with an artificial neural network comprising medical documents to accelerate positional cloning. (2009) *Nucleic Acids Res.* 37 (Web Server issue):W147-52.
2. Matsushima A., Kobayashi N., Mochizuki Y., Ishii M., Kawaguchi S., Endo T.A., Umetsu R., Makita Y., Toyoda T. OmicBrowse: a Flash-based high-performance graphics interface for genomic resources. *Nucleic Acids Res.* 37(Web Server issue):W57-62.
3. Makita Y., Kobayashi N., Mochizuki Y., Yoshida Y., Asano S., Heida N., Deshpande M., Bhatia R., Matsushima A., Ishii M., Kawaguchi S., Iida K., Hanada K., Kuromori T., Seki M., Shinozaki K., Toyoda T. PosMed-plus: an Intelligent Search Engine that Inferentially Integrates Cross-species Information Resources for Molecular Breeding of Plants. (2009) *Plant Cell Physiol.* (in press)
4. Iida, K., Fukami-Kobayashi, K., Toyoda, A., Sakaki, Y., Kobayashi, M., Seki, M. and Shinozaki, K. (2009) Analysis of Multiple Occurrences of Alternative Splicing Events in *Arabidopsis thaliana* Using Novel Sequenced Full-Length cDNAs. *DNA Res.*
5. Iida, K., Jin, H. and Zhu, J. (2009) Bioinformatics analysis suggests base modifications of tRNAs and miRNAs in *Arabidopsis thaliana*. *BMC Genomics*, 10, 155.
6. Endoh, M., Endo, T., Endoh, T., Fujimura, Y., Ohara, O., Toyoda, T., Otte, A., Okano, M., Brockdorff, N., Vidal, M. et al. (2008) Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. *Development*, 135, 1513-1524.
7. Kaminuma, E., Masuya, H., Miura, I., Motegi, H., Takahasi, K., Nakazawa, M., Matsui, M., Gondo, Y., Noda, T., Shiroishi, T. et al. (2008) Objective evaluation measures of genetic marker selection in large-scale SNP genotyping. *J Bioinform Comput Biol*, 6, 905-917.
8. Kaminuma, E., Yoshizumi, T., Wada, T., Matsui, M. and Toyoda, T. (2008) Quantitative analysis of heterogeneous spatial distribution of *Arabidopsis* leaf trichomes using micro X-ray computed tomography. *Plant J*, 56, 470-482.
9. Kim, J., To, T., Ishida, J., Morosawa, T., Kawashima, M., Matsui, A., Toyoda, T., Kimura, H., Shinozaki, K. and Seki, M. (2008) Alterations of lysine modifications on the histone H3 N-tail under drought stress conditions in *Arabidopsis thaliana*. *Plant Cell Physiol*, 49, 1580-1588.
10. Kurihara, Y., Matsui, A., Kawashima, M., Kaminuma, E., Ishida, J., Morosawa, T., Mochizuki, Y., Kobayashi, N., Toyoda, T., Shinozaki, K. et al. (2008) Identification of the candidate genes regulated by RNA-directed DNA methylation in *Arabidopsis*. *Biochem Biophys Res Commun*, 376, 553-557.
11. Matsui, A., Ishida, J., Morosawa, T., Mochizuki, Y., Kaminuma, E., Endo, T., Okamoto, M., Nambara, E., Nakajima, M., Kawashima, M. et al. (2008) *Arabidopsis* transcriptome analysis under drought, cold, high-salinity and ABA treatment conditions using a tiling array. *Plant Cell Physiol*, 49, 1135-1149.
- 【和文誌】 1. 豊田 哲郎 創薬に向けた生命情報基盤の構築(2009)遺伝子医学M00K「次世代創薬テクノロジー・実践：インシリコ創薬の最前線2009」14号
2. 豊田 哲郎 ポジショナルキャンディデートクローニングのための統合データベース(2008)医学のあゆみ 225(9) 770-774
3. 花田 耕介, 川口 修治, 豊田 哲郎, 篠崎 一雄 mRNAの転写活性を示すncRNA遺伝子の網羅的な探索 日本化学会情報化学部会誌(2008)26(3)62-65.

Information

独立行政法人理化学研究所 横浜研究所 生命情報基盤研究部門

〒230-0045 神奈川県横浜市鶴見区末広町1-7-22 東研究棟E219

URL: <http://www.base.riken.jp/> <http://database.riken.jp/> <http://omicspace.riken.jp/>