

理研生命情報(公開)基盤研究部門

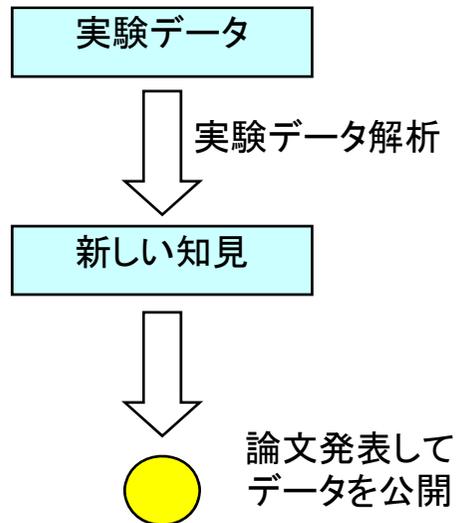
平成20年4月3日

独立行政法人理化学研究所

豊田哲郎

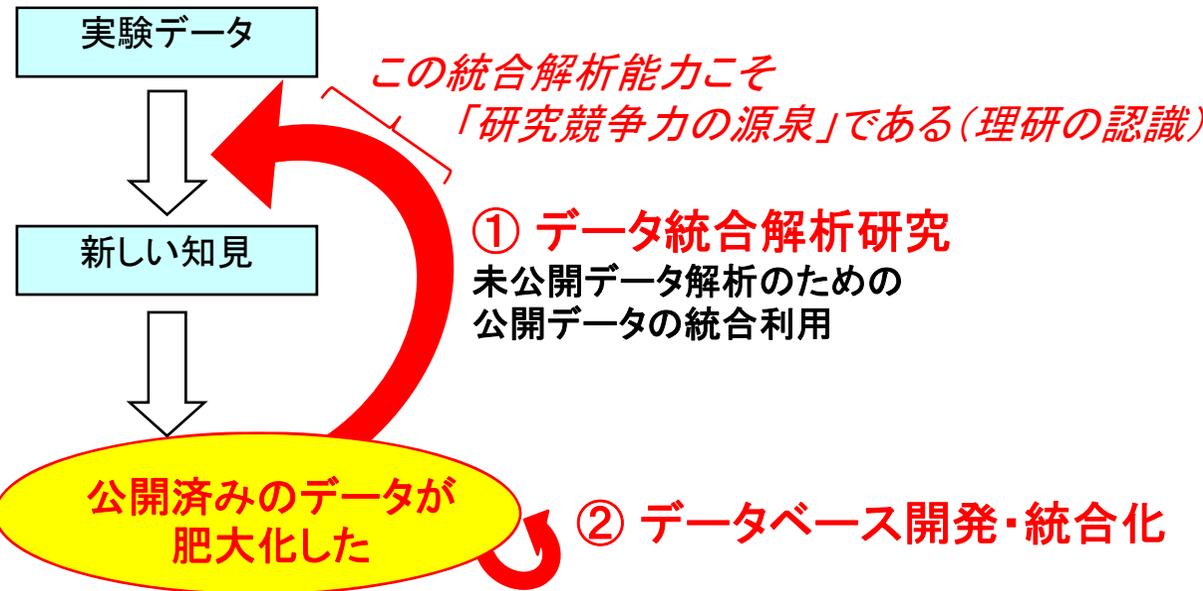
バイオインフォマティクス = 研究競争力の源泉

以前の解析フロー



現在の解析フロー

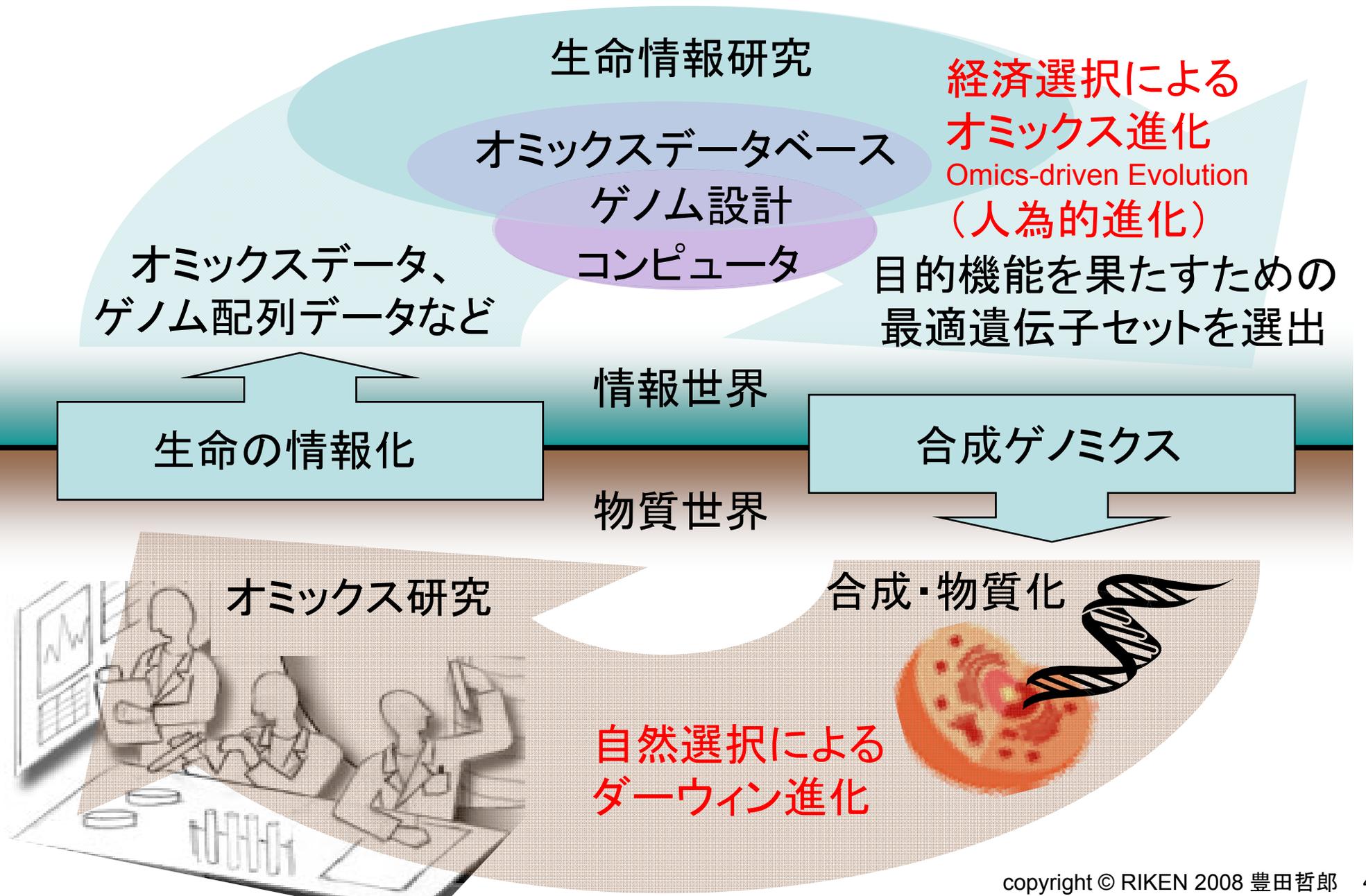
新たな実験データの解析には
既存公開データとの統合解析が不可欠になった



③今後の課題： 大規模データの半自動処理ラインが、最終的な付加価値を生み出すステップに不可欠

	自動車産業	ライフサイエンス
<p>本体が持つべき機能 研究競争力の源泉</p> <p>(最終的な付加価値の生産)</p>	<p>ニーズ変化に即時対応可能な半自動的組立て生産ライン</p> 	<p>ニーズ変化に即時対応可能な半自動的情報処理ライン</p> <p>大規模データの自動処理技術 (ラボ内の自動処理ライン)</p> <p>↓ 新規／既存データの統合解析</p> <p>↓ 研究員の知的活動による論文生産</p> <p>研究者の移動リスク対策必須</p>
<p>外注化可能な機能 コストダウン競争</p>	<p>個々の部品生産</p> <p>部品、パーツの生産は下請けの工場に発注</p>	<p>個々のオミックスデータ生産</p> <p>アウトソース化</p> 

④ 将来トレンド：“情報資源”から“新しい生物資源”を創造(例、iPS細胞)



RACの提言：バイオインフォマティクスとデータベース基盤の強化

バイオインフォマティクスの専門家を増員するとともに、簡単にアクセスできる理研全体の統合データベース、及び共通のデータスタンダード、データ公開方針を作り、理研のバイオインフォマティクスを強化すること。

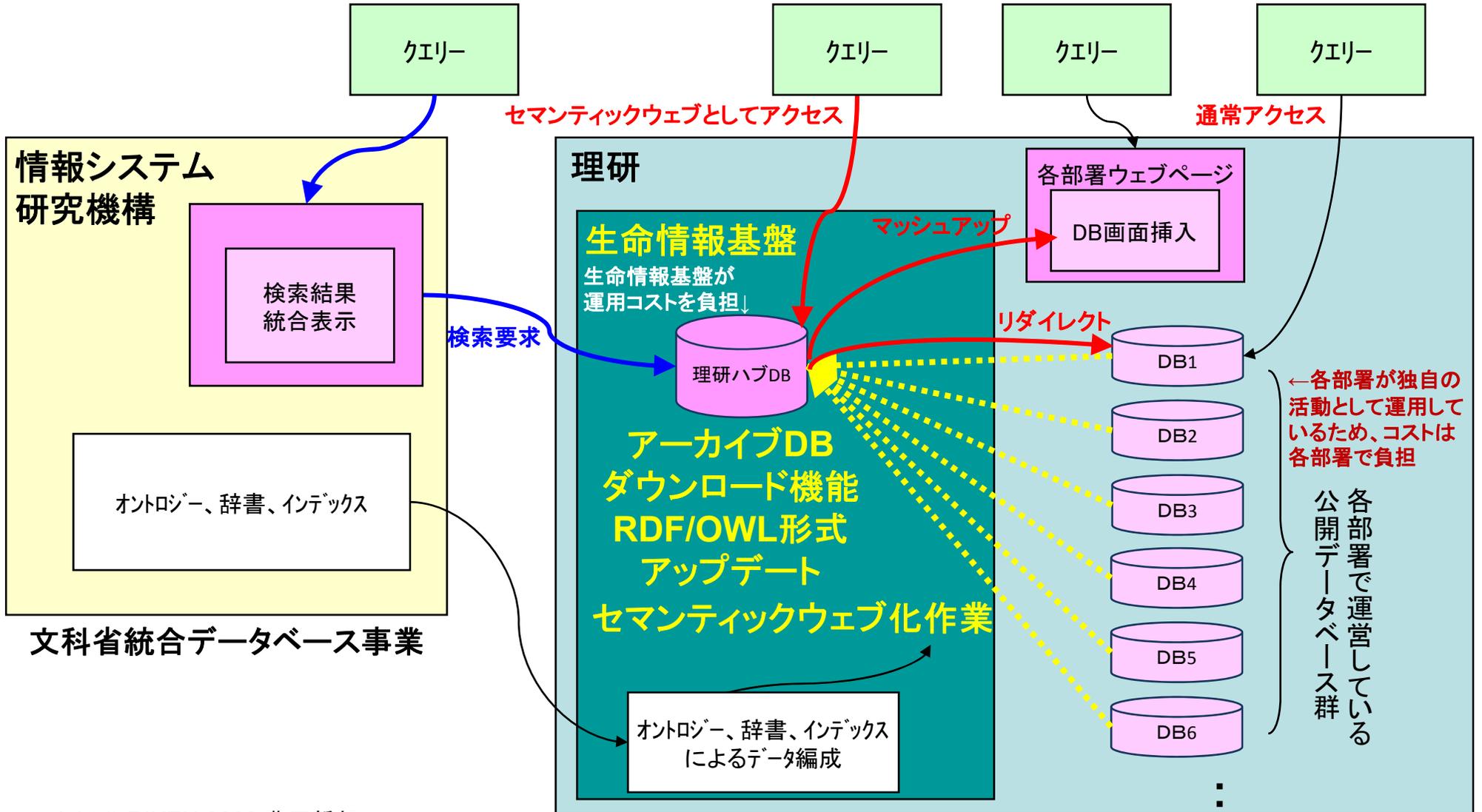
Strengthen bioinformatics at RIKEN by recruiting more specialists and by creating integrated, easily accessible RIKEN-wide databases as well as common data standards, data release policies.



【野依イニシアティブ「世界的なブランド力のある理研」に向けて】

- ◎公開データ資源における理研ブランドをインターネット上で確立
- ◎理研の公開データが利用されれば、理研の論文の引用回数も増加
- ◎正の連鎖：データ相互利用の強化 → データ解析能力の強化 → 研究競争力強化

理研データのセマンティックウェブ化を推進



セマンティックウェブ化された各種データベース

トランスポゾン挿入シロイヌナズナ変異体の表現型データベース

代謝物データベース (KNAPSAcK)

RIKEN Arabidopsis Phenome Information Database

Database

RIKEN Ds transposon line code

ID	Term(English)	Definition	Instance
12-2589-1	RIKEN Ds transposon line code		12-2589-1

Instance

12-2589-1

Definition: 日本語定義

Term(English): RIKEN Ds transposon line code

Definition: 日本語定義

Instance: 12-2589-1

Image(s):

File(s):

Class: **has_transposon insertion region**

Property: **has_transposon insertion position**

MapWord: **6000**

WebLink(s): **331**

KNAPSAcK Database

Database

KNAPSAcK record

ID	Term(English)	Definition	Instance
C00000388	KNAPSAcK record	Information on natural products has been amassed within KNAPSAcK with special emphasis on their biological origins.	C00000388

Instance

C00000388

Definition: 日本語定義

Term(English): KNAPSAcK record

Definition: Information on natural products has been amassed within KNAPSAcK with special emphasis on their biological origins.

Instance: C00000388

Image(s):

File(s):

Class: **has_molecular formula**

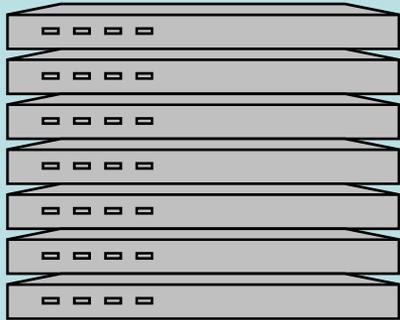
Property: **has_molecular weight**

MapWord: **304**

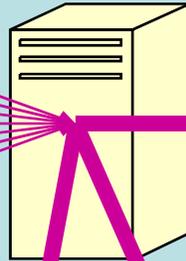
WebLink(s): **2402302654**

数百種類のゲノムアノテーションの分散共有システム (多数の外部共同研究者ごとのアクセス制御)

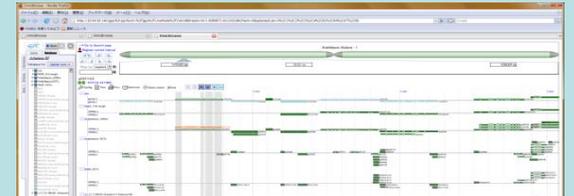
生命情報基盤研究部門



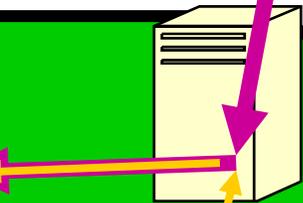
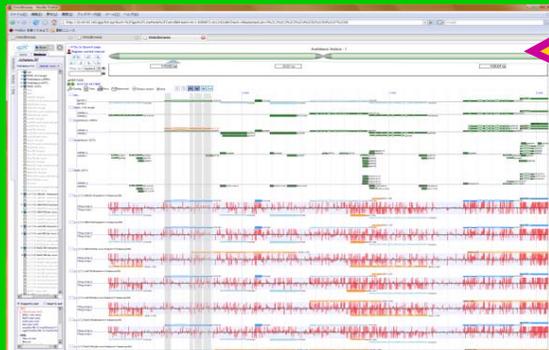
Accumulated
Public Annotations



Internet
Server



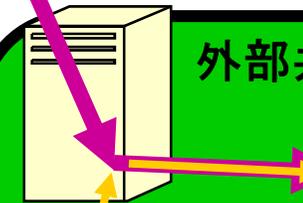
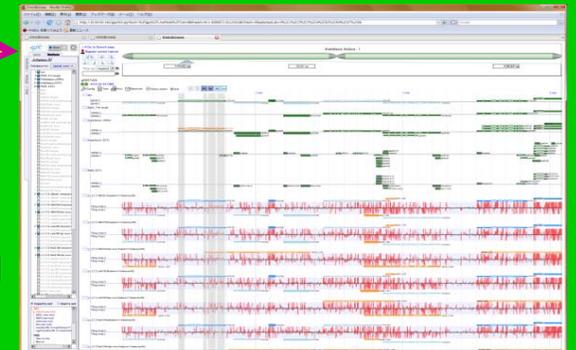
各センターの研究室



Intranet
Server

Integrating
Unpublished Data

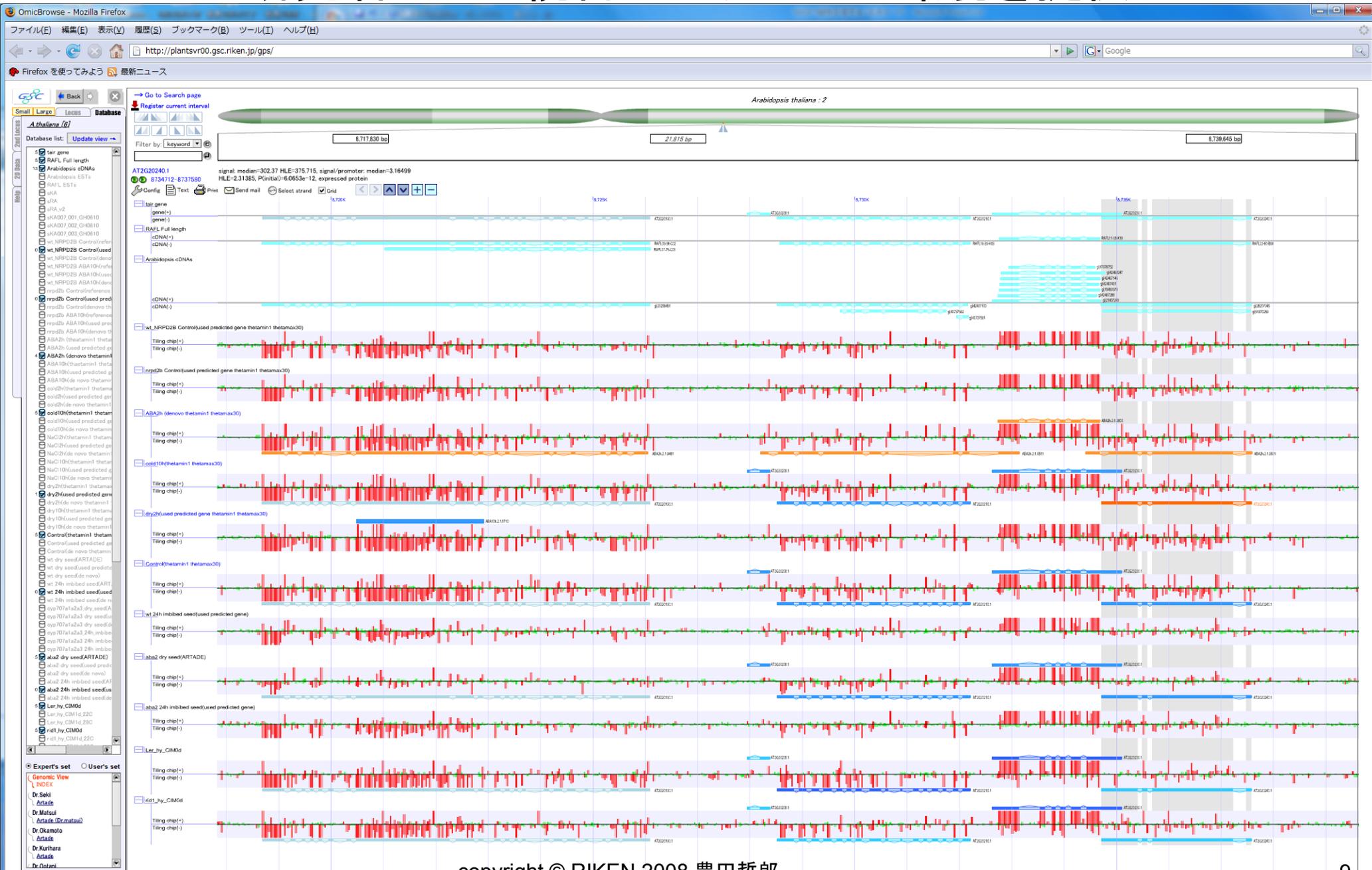
外部共同研究者



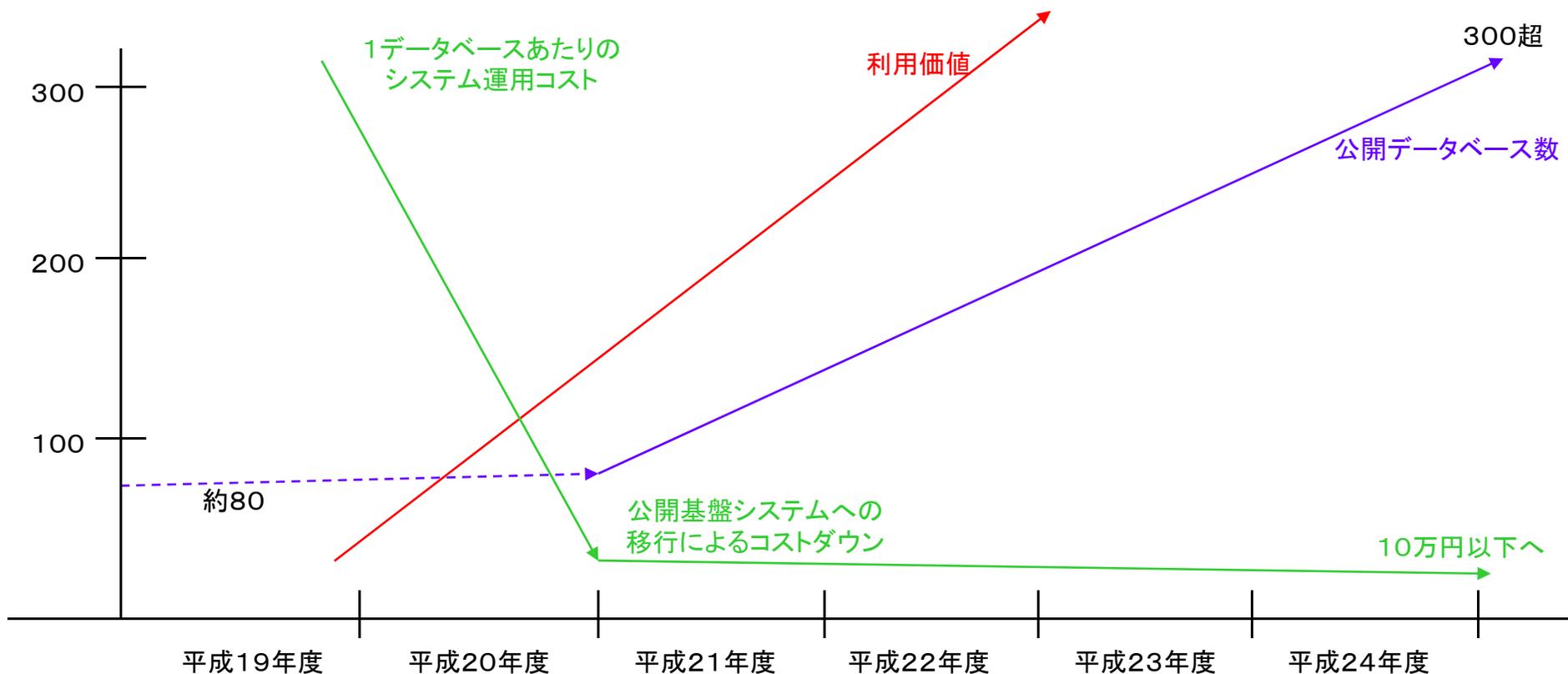
Intranet
Server

Integrating
Unpublished Data

研究者ごとに統合データベース環境を提供



公開データベースの費用対効果 今後5年間の見通し



生命情報基盤部門からの貢献

◎生命情報公開基盤の構築により、皆がシステム運用の負担から解放されることで、

- 誰でも容易にデータ公開できる環境を実現する
- コピーレフトでのデータベース共有を推奨する
- 国際的なデータ連携における情報基盤を提供する
- 国内外のデータベース統合化事業に協力する
- アーカイブ化で公開データの永続性を保障する
- データの帰属先(著作権やライセンス等)をきちんと示すことで提供者の利益を保護する

◎目標:理研からのデータベース公開活動を5年間で3倍に増やす

◎理研データベースのセマンティックウェブ化
膨大なデータの半自動解析フローを実現

◎バイオインフォマティクス研究
オミックスデータの統合解析(ウェットラボとの共同研究)

理研横浜研究所の新組織

【平成18～19年度】

次期中期計画【平成20～24年度】

