

BioHackathon 2008の報告

片山俊明 <ktym@hgc.jp>

東京大学医科学研究所ヒトゲノム解析センター
ゲノムデータベース分野 助教

「生物情報資源の相互運用性」研究会@遺伝研 2008/3/25

BioHackathon 2008

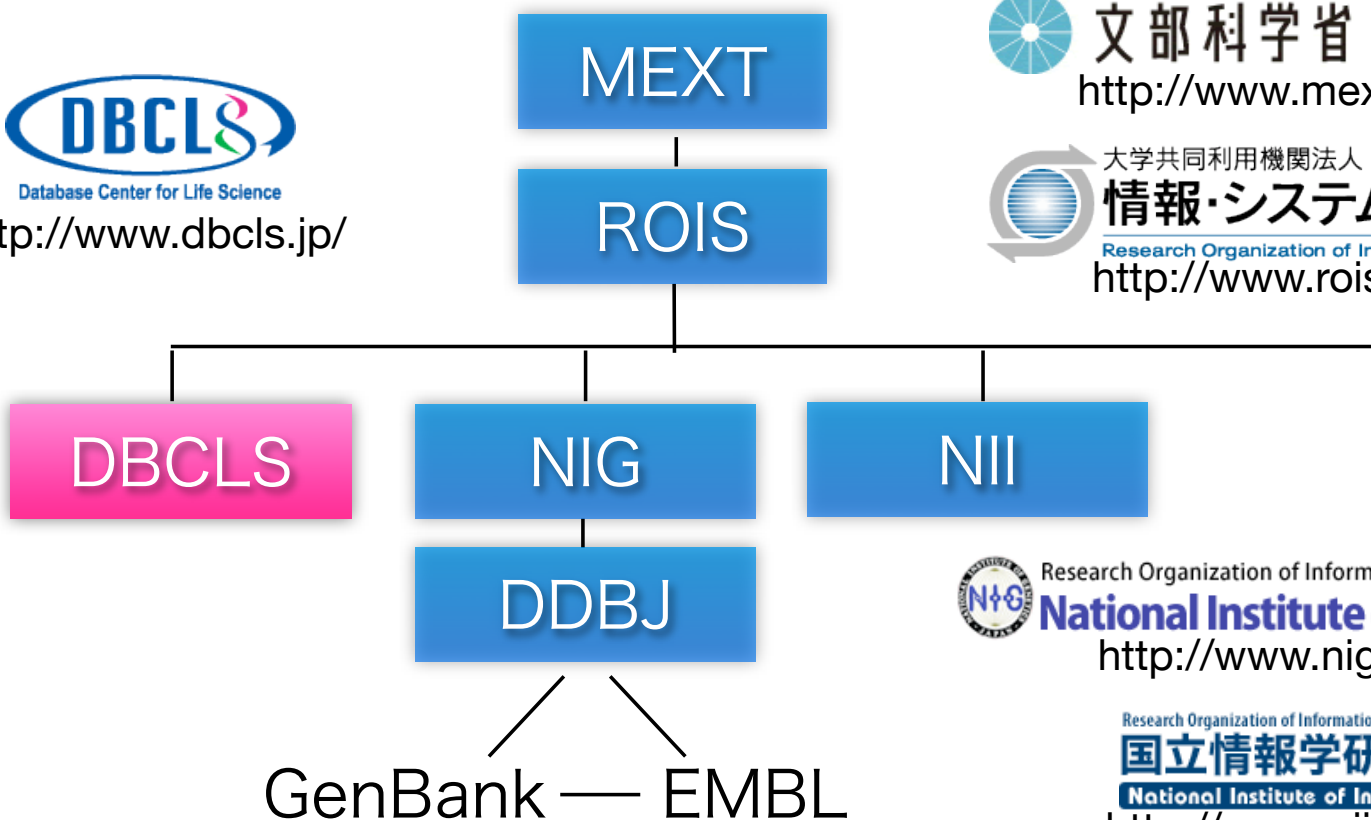


- ・ 2008/2/11-15 に東京で開催された開発主体の国際会議
- ・ ライフサイエンス統合データベースセンター + CBRC 主催
- ・ 今回のテーマはウェブサービス
- ・ 目的は国内外のサービスを統合するために必要な現状調査と技術開発



DBCLS - スポンサー

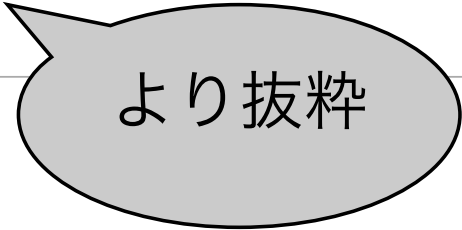
・ ライフサイエンス統合データベースセンター



:

ライフサイエンス統合データベースセンター
共通基盤技術開発チーム
2007年度活動計画

DBCLS: 高木, 山口, 山本
かずさ: 中尾
HGC: 川島, 片山
+ 辻井研, 松本研, CBRC

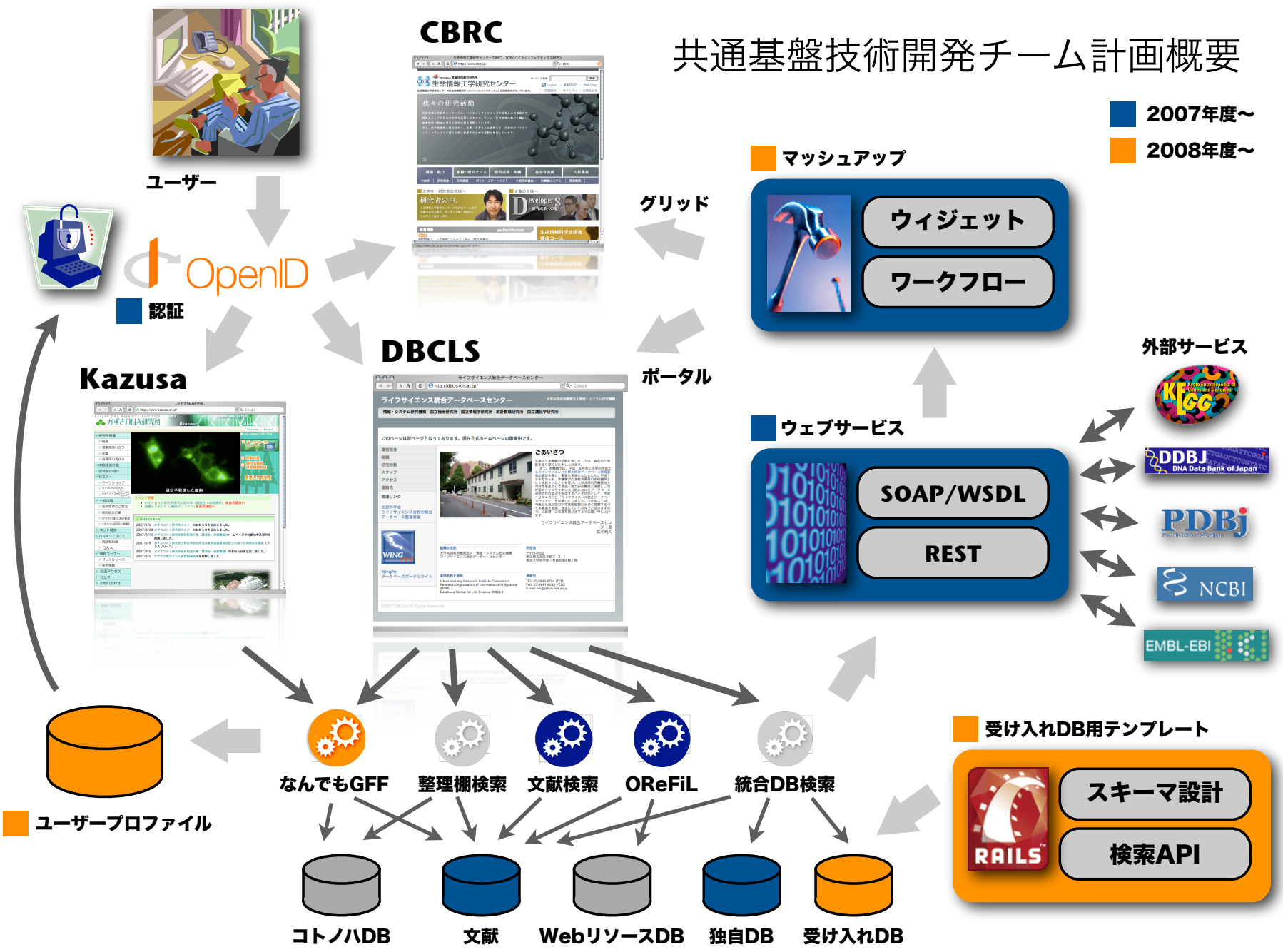


より抜粋

2007/9/12@DBCLS

共通基盤技術開発チーム計画概要

■ 2007年度～
■ 2008年度～



基盤整備の4本柱

- 次世代検索技術開発

1. 文献検索

- 受け入れDB支援

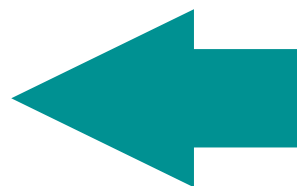
2. DB構築テンプレート

- サービス統合支援

3. 統合ウェブサービス

- ポータル連携支援

4. OpenID ユーザ認証



● 3. 統合ウェブサービス

- ・ 文献検索など、統合DB独自開発ツールのウェブサービス化 → ウィジェット
- ・ 受け入れDBのウェブサービス対応による利用率向上を図る → ワークフロー
- ・ 国内外のウェブサービスを統合し SOAP/WSDL + REST で提供
 - ・ KEGG, DDBJ, PDBj, Riken, CBRC など + EBI, NCBI などに対応

- ・ 課題：サーバ毎に使い方や命名規則が異なる
 - ・ 統一的なユーザーインターフェイス（メソッド名、引数）を整備 -- プロキシ的メタサーバ
- ・ 課題：多くのサーバは使い方がよく分からない
 - ・ 統合DBのサービスで全機能を吸収し、分かりやすいドキュメントを整備
- ・ 課題：サービス間でデータの受け渡しが出来ない
 - ・ 共通のデータフォーマットを整備しワークフローの構築を実現 -- BioMOBY, Taverna
- ・ 課題：サービスが必ずしも安定ではない
 - ・ 統合DBのサービスでタイムアウトやエラーの検出などを行い、使い勝手を向上

基盤チームと他チームとの連携

- ヒト統合DBチーム

様々なデータの有機的な統合と文献検索技術開発による支援

- DB受け入れチーム

DBの受け入れから統合的な検索サービス公開までの技術的な支援

- 教育チーム

必要なツールやワークフローのウェブサービスによる提供

- ポータルチーム

魅力的なポータルサイト構築のための提案と技術的な支援

基盤チームと協力研究室との連携

- ・ CBRC

- 課題：CBRC のリソースを中心としたワークフロー構築

- ・ ウェブサービスの相互提供によるワークフロー構築支援

- ・ ポータルチームとの連携

BioHackathon 2008 開催の意図

- ・ ライフサイエンス統合データベースセンターのミッション
 - ・ 国内のライフサイエンスDBを統合する
 - ・ 現実には（特にDDBJやKEGGなど大規模なDBの統合は） 困難
 - ・ バーチャルに統合するのが現実的
 - ・ 現状ではおそらくウェブサービスが最適



- ・ どのように統合するのがベストか
 - ・ BioMoby?
 - ・ WS-I?
 - ・ セマンティックウェブ?
 - ・ 既存のライブラリ?
 - ・ 瀧内メソッドの出番

新しい技術を習得することは難しい

まず最初に直面する状態・・・自分の知らない領域で身動きが取れない状態・・・

ここから抜け出す事が技術自体の学習よりも遥に困難なのだ

その分野に精通した人が適切な入門書(ドキュメント)を提示してくれる・・・

さらに言えば、出向いてもらい、手取り足取りボランティアで教えてもらう・・・

私はそんな願望一杯の状況を自分で簡単に作り出す恐るべき方法を思いついた

「勉強会」にして人を集めればいいんだ・・・

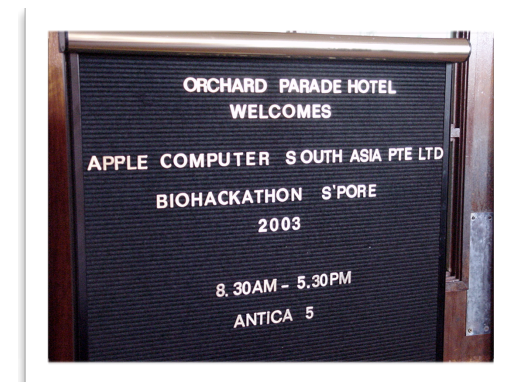
「序章：頼むより巻き込め！！」(瀧内メソッド入門)より抜粋

そもそもBioHackathonとは

- ・ Hackathon = ハック + マラソン = 開発合宿
 - ・ 約1週間、20~40人くらいの開発者を缶詰にしプログラム開発
- ・ BioHackathon
 - ・ バイオインフォのツール、ライブラリに様々な不整合や課題があった
 - ・ 世界中に分散している BioPerl など Open Bio* の開発者を集めて改善
 - ・ バイオインフォに興味をもっていた O'Reilly などがスポンサーに
 - ・ オープンソースによる開発の利点とコミュニティの醸成
 - ・ 最先端、最高レベルの開発者を1週間タダ働きさせられる
 - ・ 開発者間での情報流通によるお互いのフィードバックと効率
 - ・ 成果物は誰でも利用できるように公開される
 - ・ 結果的にバイオインフォの環境が改善され市場も開拓される？！

BioHackathonという文化

- ・ 2002 第1回, 第2回 BioHackathon
 - ・ Open Bio* の各プロジェクトからコアディベロッパーが集結
 - ・ 1月 アリゾナ州 Tucson - スポンサー O'Reilly
 - ・ 2月 南アフリカ Cape Town - スポンサー Electric Genetics
- ・ 2003 第3回 BioHackathon
 - ・ 2月 シンガポール - スポンサー Apple Asia
 - ・ 主に OBDA の整備が完了
 - ・ BioSQL, Chado, GBrowse, BioMoby など
- ・ BioPAX, GMOD, SBML などさまざまな Hackathon
- ・ 2006 Phyloinformatics Hackathon
 - ・ 12月 ノースキャロライナ州 - スポンサー NESCent, Duke大学



BIOHACKATHON 26-28 JAN 2002, TUCSON, ARIZONA, USA



BioHackathon #2

- Cape Town, South Africa
- 2002/2/24-3/1

SAVE

MASTER Module LIST

	Spec	Java	Perl	Python	Ruby	192	168	0.95
Registry	✓	Brian	✓	Jeff				
Index-BDB	✓		Lincoln	Andrew				
Index-Flat	✓	Matt	Module	Andrew				
BioFetch	✓	Defflat	Kracker	Andrew	server	Antonia		
CORBA	✓	ch ✓ swie ✓	ch ✓ swie ✓	ch ✓ swie ✓				
BioSQL	✓	Paul ✓ (with) ✓ wite ✓	Paul ✓ Bik ✓ Mark ✓	Paul ✓ Rand ✓ M ✓				
Xembl	✓	Cyl ✓ SOAP	gi ✓ SOAP	gi ✓ SOAP				
UPDI		Matt ✓ Brian	soap ✓	soap X				

REGISTRY

192 168 0.95

http://

int

key

Seco

P

P

Gonbank/idea

37

14 + 1:24 minutes → Plat

OBDA

- Open Bio* Database Access
 - BioRegistry
 - BioFlat
 - BioFetch
 - BioSQL

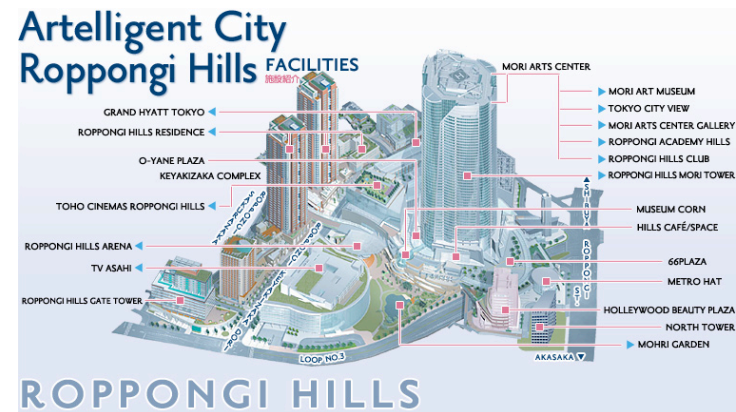
BioHackathon 2008 - program

• 2/11 morning session - presentations

- From Web API for Biology (WABI) to Semantic Web API for Biology (SABI) - Hideaki Sugawara
- Current status of the BioMOBY project and vision for the future directions - Mark Wilkinson
- The EMBRACE project and WS-I standard - Jan Christian Bryne
- Soaplab2 project to wrap up command line packages - Martin Senger
- Proxying legacy applications and CGIs into BioMOBY - Paul Gordon
- Generation Challenge Program effort at building interoperability - Richard Bruskiwich

• 2/11 afternoon session - open space

- ServiceProvider_workgroup
- OpenBio_workgroup
- Workflow_workgroup
- BioMOBY_workgroup
- Exchangeformat_workgroup
- Semanticweb_workgroup
- PhyloWS_workgroup
- Distributed storage - initiative (BIO.SLURP)
- Distributed services - initiative (BIO.ORG.ANISM)



• 2/12-15 hack, hack, hack

• We will have banquet on 2/12 18:30- @CBRC

分野間の連携

ウェブサービスの
プロバイダ

Open Bio* の
開発者

ウェブサービスの
クライアント開発者

BioMoby

交換データ
フォーマットの
標準化

参加者の内訳（概算、のべ数）

List of Web Service Hackathon attendees

Fulltime participants

- Organizers
 - Toshiaki Katayama (⇒ Univ of Tokyo, Japan; ⇒ BioRuby, ⇒ KEGG API, ⇒ KEGG DAS)
 - Yasunori Yamamoto (⇒ DBCLS, Japan; Text mining)
 - Hong-Woo Chun (DBCLS, Japan)
 - Atsuko Yamaguchi (DBCLS, Japan)
 - Tamotsu Noguchi (⇒ CBRC, Japan)
 - Kiyoshi Asai (CBRC, Japan)
 - Toshihisa Takagi (DBCLS, Japan)
- Web service providers
 - DDBJ
 - Hideaki Sugawara (⇒ DDBJ, Japan; ⇒ DDBJ XML)
 - Yasumasa Shigemoto (DDBJ, Japan; DDBJ XML)
 - Yoshikazu Kuwana (DDBJ, Japan; DDBJ XML)
 - KEGG
 - Shuichi Kawashima (Univ of Tokyo, Japan; KEGG API)
 - PDBj
 - Akira Kinjo (Osaka Univ, Japan; ⇒ PDBj; ⇒ PDBj REST API)
 - Daron Standley (Osaka Univ, Japan; ⇒ PDBj)
 - CBRC
 - Toshiyuki Tashiro (CBRC, Japan)
 - Tatsuya Nishizawa (IMS/DBCLS/CBRC, Japan; PDBj)
 - Hiroyuki Sato (IMS/DBCLS/CBRC, Japan)
 - Oto Maekawa (IMS/DBCLS/CBRC, Japan)
 - EBI
 - Richard Cote (EBI, UK; PICR and OLS web services)
 - Florian Reisinger (EBI, UK; ENVISION/ECORE, Ensembl, UniProt-DAS, ArrayExpress, PRIDE, IntAct, and Reactome)
 - Arnaud Kerhornou (EBI, UK; ex-INB, IMIM)
 - Other web service providers and standard exchange format developers
 - Evangelos Pafilis (European Molecular Biology Laboratory; SRS web services)
 - Jan Christian Bryne (Computational Biology Unit, Bergen; EMBRACE)
 - Matthew Pocock (Newcastle, UK; Taverna, BioJava)
 - Oswaldo Trelles (INB, Spain; MOWServ)
 - José M. Fernández (INB, Spain; IWWE&M, BioMOBY, MOWServ)
 - Keiichiro Ono (⇒ UCSD, US; ⇒ Cytoscape)
 - Bruno Aranda (EBI, UK; ⇒ IntAct)
 - Lukasz Salwinski (UCLA, US; ⇒ DIP)
 - Michael Kuhn (EMBL, Germany; STRING)
 - Ken Fukuda (CBRC, Japan; BioPAX, INOCH)
 - Akira Funahashi (Keio Univ, Japan; SBML)
 - Rene Ranzinger (German Cancer Research Center, Germany; Glycoinformatics, EUROCarbDB, GlycomeDB)
 - Will York (Georgia Univ, US; Glycoinformatics, GLYDE)
 - Kiyoko Kinoshita (Soka Univ, Japan; Glycoinformatics, RINGS)
 - Chikashi Nobata (Manchester Univ, UK; Text mining)
 - Yoshinobu Kano (Tokyo Univ, Japan; Text mining)
 - Rutger Vos (University of British Columbia, Canada; ⇒ CIPRES, ⇒ Nexml, ⇒ EvoInfo)
- Open Bio* developers
 - BioRuby
 - Mitsuteru Nakao (Kazusa DNA Res, Japan; BioRuby)
 - Naohisa Goto (Osaka Univ, Japan; BioRuby)
 - Jan Aerts (Roslin Institute, UK; BioRuby)
 - Piotr Prins (Netherlands; BioRuby; ⇒ Wageningen University; ⇒ RUG Bioinformatics)
 - Alex Gutteridge (UK; BioRuby)
 - Christian Zmasek (Burnham Inst, US; BioRuby)
 - Raoul Jean Pierre Bonnal (CNR-ITB, Italy; BioRuby, BioSQL)
 - BioJava
 - Mark Schreiber (Novartis Institute of Tropical Diseases, Singapore; BioJava)
 - Richard Holland (EBI, UK; BioJava, BioMart)

- DBCLS - 5名
- DDBJ - 2名
- KEGG - 2名
- PDBj - 2名
- CBRC - 4名
- EBI - 3名
- その他のWSプロバイダ - 9名
- データ型 - 8名
- テキストマイニング - 3名
- クライアント開発者 - 6名
- BioRuby - 7名
- BioJava - 2名
- BioPerl - 3名
- BioMoby - 7名



Roll call at Villa Fontaine Shiodome

2008 Feb 11

academyhills 40F
Quarter C4
キャラントC4
10:30~
17:00 Bio Hackathon



BioHackathon
2008
Feb 11-15 @ Tokyo, Japan
http://biohackathon-2008-tokyo.jp

BioHackathon
2008
Feb 11-15 @ Tokyo, Japan
http://biohackathon-2008-tokyo.jp

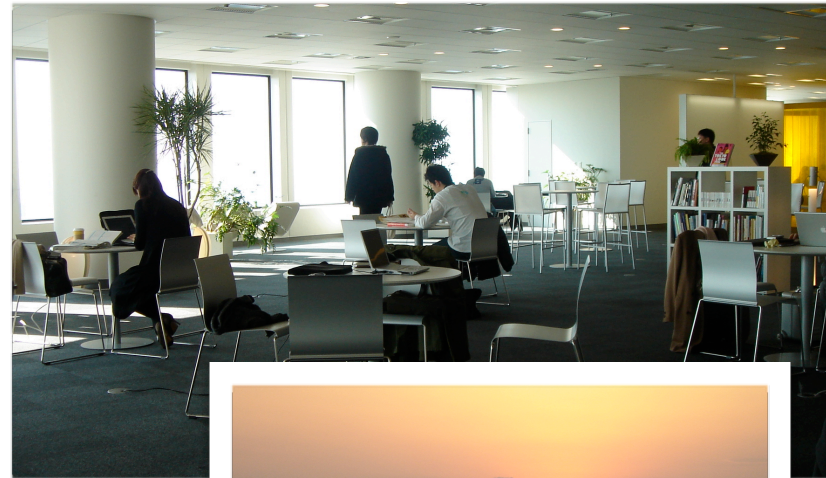
BioHackaton 2008 Open Space topics

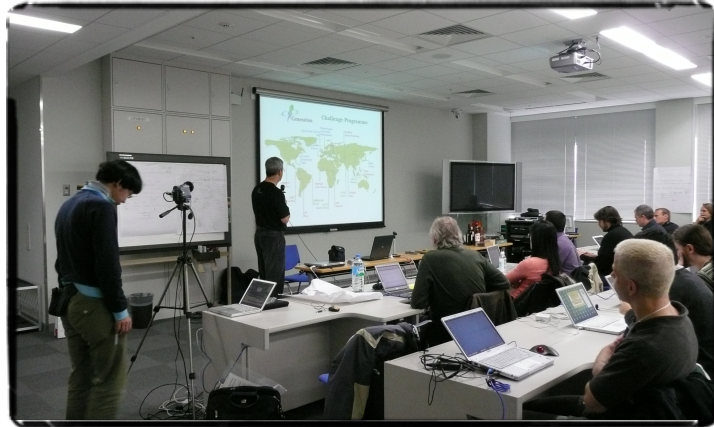
- ・ オントロジー - Richard Bruskiwich
 - ・ データ型、サービスの分類
- ・ 交換データの標準化
 - ・ Glycoinformatics (糖鎖) - Will York
 - ・ 相互作用ネットワーク, DAS - Bruno Aranda
 - ・ テキストマイニング - Yoshinobu Kano
 - ・ Phyloinformatics (系統解析) - Hilmar Lapp
- ・ Async service (非同期通信) - Jose, Fernandez
 - ・ WSRF?
- ・ Large data (大規模データ) - Tom Oinn
 - ・ SOAP attachments
- ・ セキュリティ - Oswaldo Trelles
- ・ サービスの Description & Discovery - Mark Wilkinson
- ・ Open Bio* 開発グループ (BioPerl, BioRuby, BioJava)
 - ・ BioSQL - Mark Schreiber

- ・ 相互運用性のあるワークフローの構築 - Tom Oinn, Stuart Owen
- ・ 相互運用性のための規格を提言する manifesto

<http://hackathon.dbcls.jp/wiki/ListOfTopics>

Open Space





Hack
Hack
Hack

&

Banquet

@

CBRC

2/12-15

2008

主要センターのウェブサービスと BioMoby

- NCBI SOAP
- EBI Web Service
- DDBJ XML (WABI)
- KEGG API
- PDBj SOAP
 - 全て独自に開発実装、特に標準化はされていない
 - 使われているデータ型もバラバラ
 - サービスの命名規則もバラバラ
 - 非同期通信などの仕様もバラバラ
- BioMoby
 - 植物ゲノム関係などを中心に数十サイトが対応
 - 共通のデータ型（ただし独自XML）
 - スペインの MOWServ などでグリッド化の実例



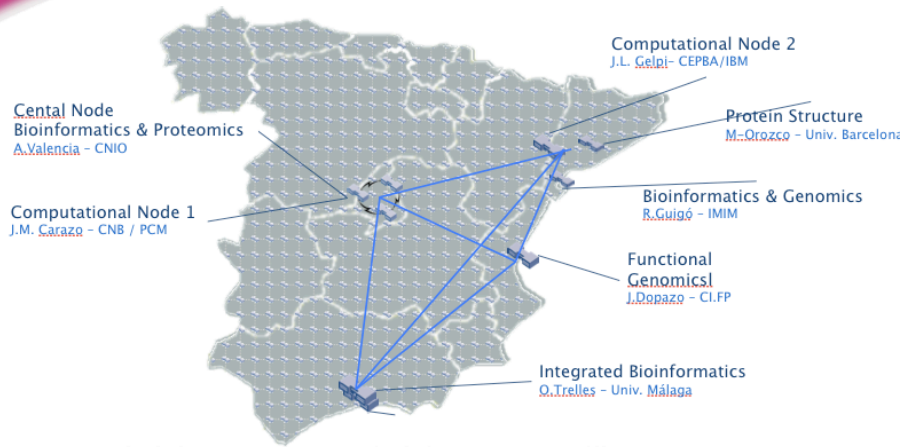
```
<Object namespace="GenBank/  
AC" ID="AY070397.1"/>
```

```
<Sequence namespace="GenBank/  
AC" ID="AY070397.1"/>  
<Length>960</Length>  
<SequenceString>  
aacaaaaagattaaacaagagag...  
</SequenceString>  
</Sequence>
```

スペインの MOWServ による統合例



National Institute for Bioinformatics
www.inab.org
Integrated Bioinformatics



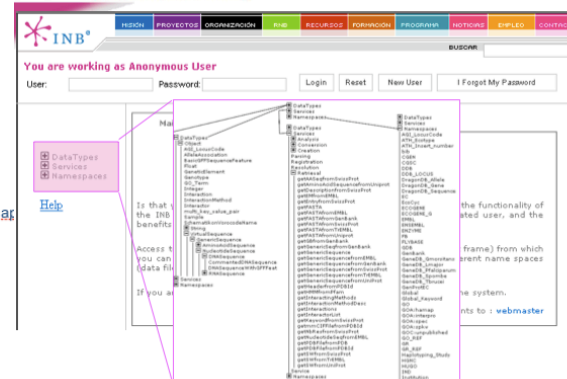
Scalable & expandable “Virtual” Organization

Integrated Bioinformatics, INB-UMA

O.Trelles, BioHackaton-Jap

Slides by Prof. Trelles (INB/MOWServ)

枠組みとして BioMoby を採用
 スペイン国内の BI センターをグリッド化
 データ型とサービスのオントロジー
 ワークフロー作成
 ジョブ管理

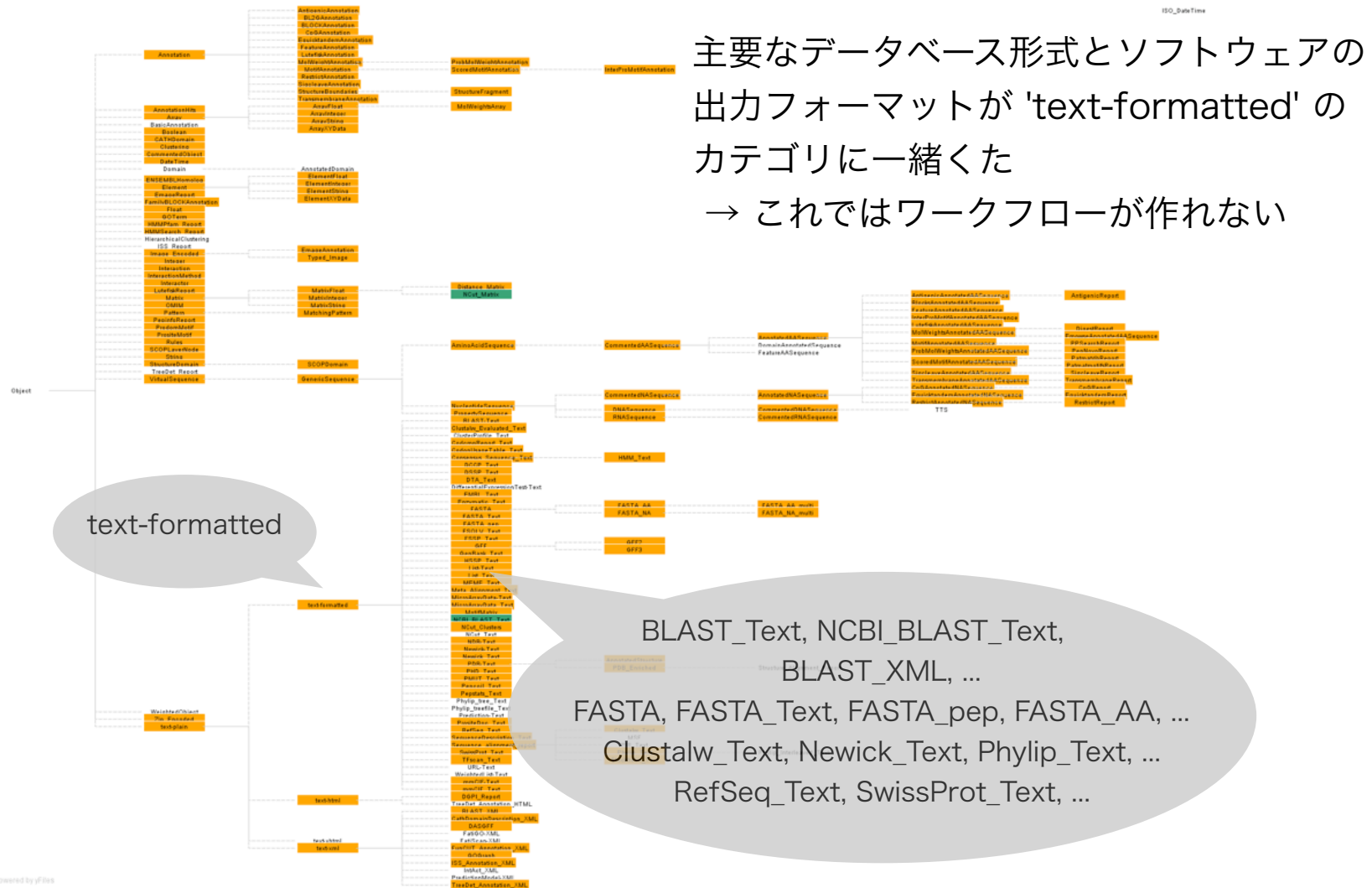


Quality control:
 registering / daily
 availability /
 documentation

Favourites section

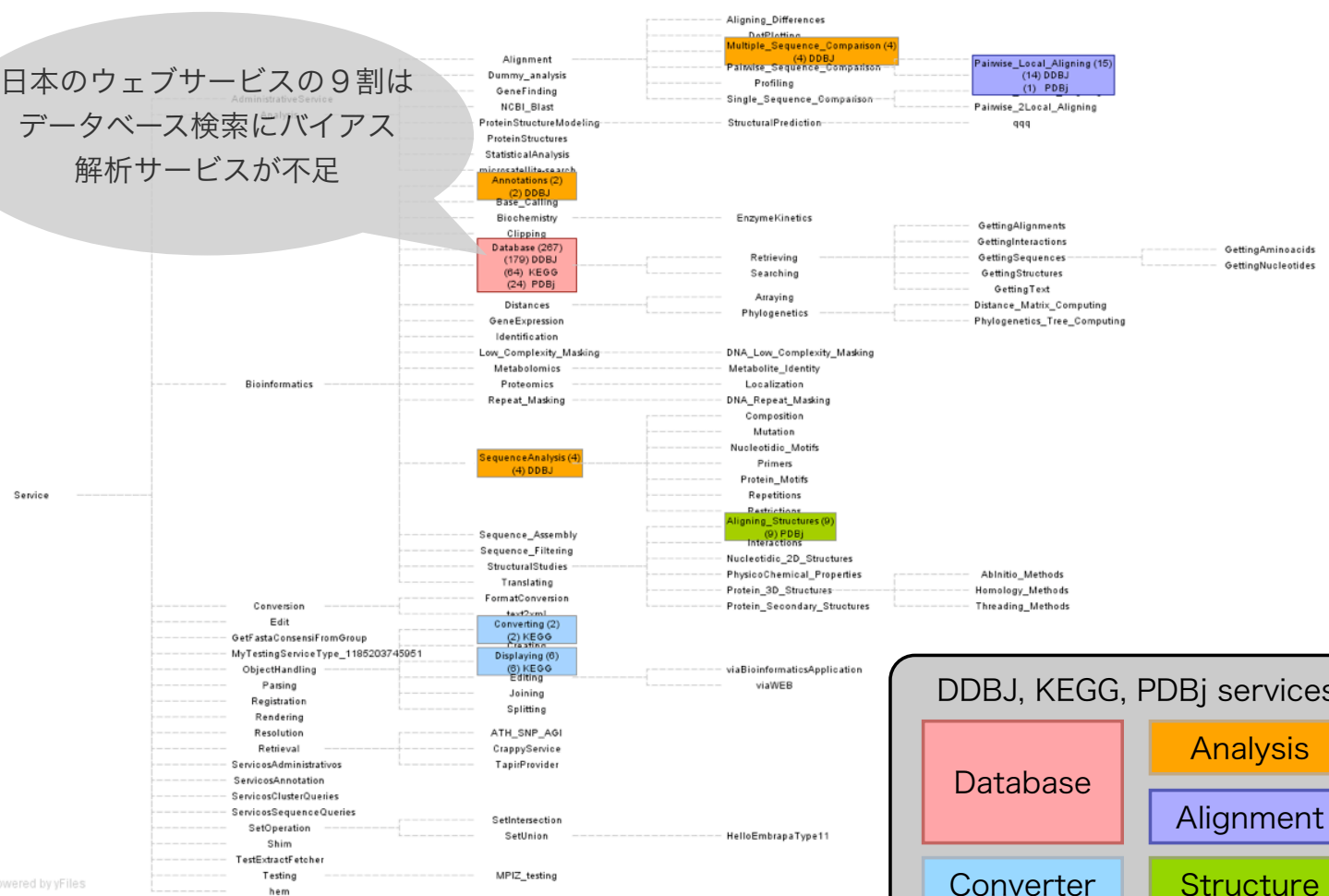
Intelligent client for integrating bioinformatics services”, Navas-Delgado, et al.
 Bioinformatics, vol.22 no.1 2006 pages 106-111

BioMoby/MOWServ object ontology



BioMoby service ontology

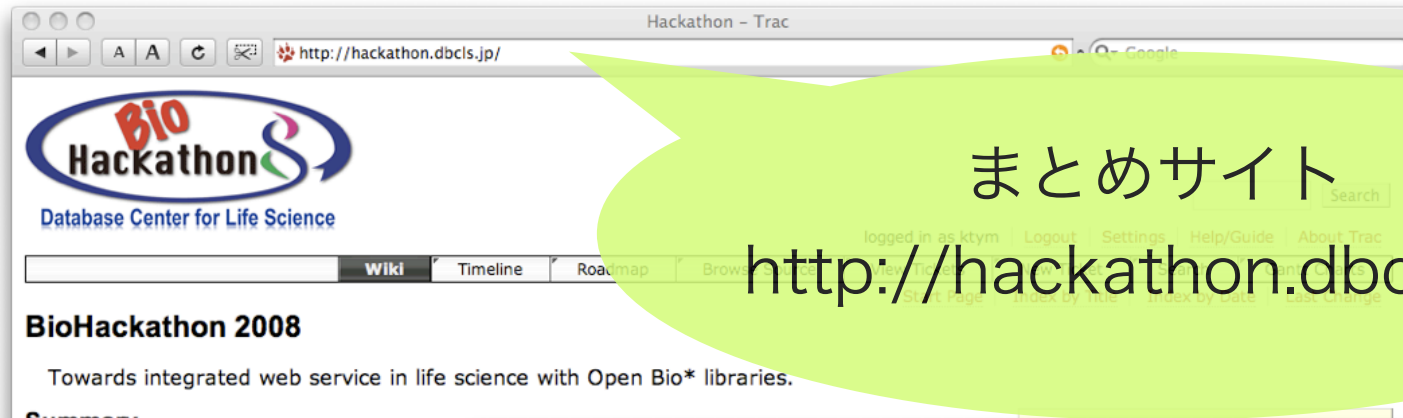
日本のウェブサービスの9割は
データベース検索にバイアス
解析サービスが不足



DDBJ, KEGG, PDBj services

Database	Analysis
Converter	Alignment
	Structure

ワークグループごとの成果



The screenshot shows the homepage of the BioHackathon website. The browser address bar displays "http://hackathon.dbcls.jp/". The page features the BioHackathon logo and the text "Database Center for Life Science". A navigation menu includes "Wiki", "Timeline", and "Roadmap". The main heading is "BioHackathon 2008" with the subtitle "Towards integrated web service in life science with Open Bio* libraries." Below this is a "Summary" section.

まとめサイト

<http://hackathon.dbcls.jp/>

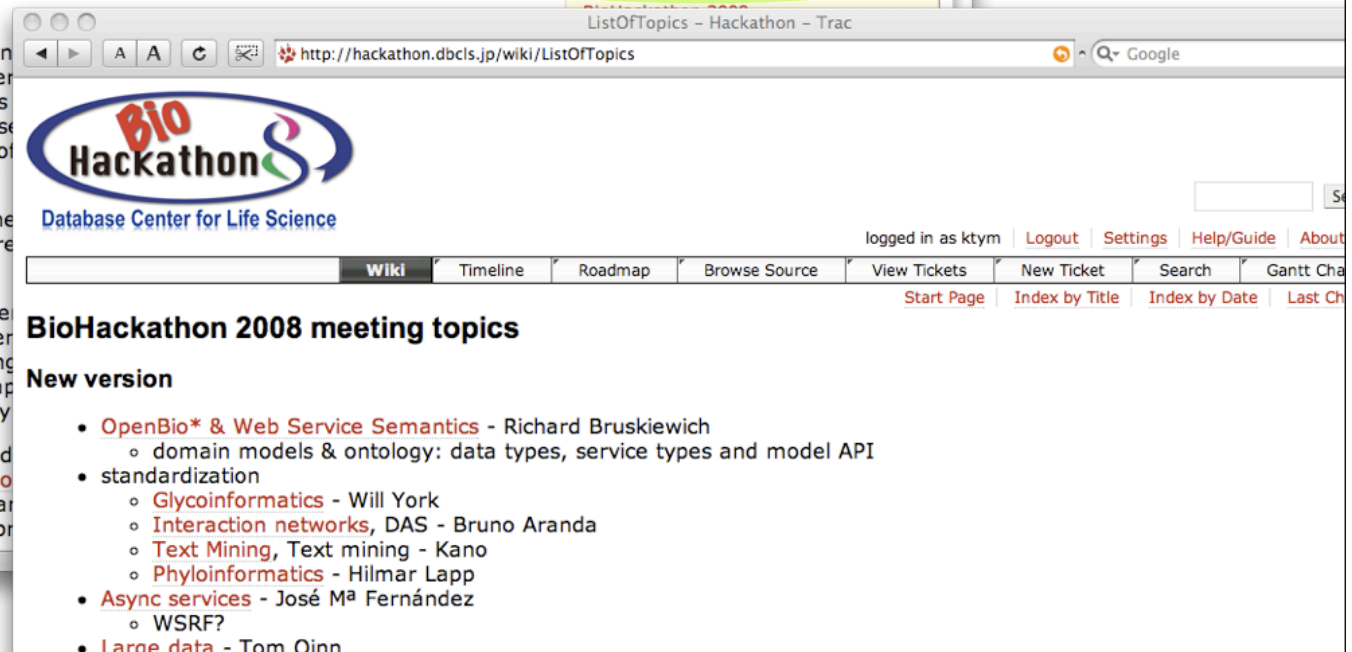
Summary

The number of web service providers in year. In theory, these services are inter languages. However, each service uses naming conventions. Moreover, these s (partly, due to the lack of compliance of library).

This lack of interoperability prevents the users and developers alike and they are research.

The ⇒ [BioMOBY](#) project has tried to ove data types and methods used in the ser providers which are not yet participating are several web service providers in Jap but none of these are integrated in any

For this reason, at ⇒ [DBCLS](#), we started service providers. If all [⇒ <http://www.o> interoperability among these services ar of web services by every web service pr



The screenshot shows the "List of Topics" page on the BioHackathon website. The browser address bar displays "http://hackathon.dbcls.jp/wiki/ListOfTopics". The page features the BioHackathon logo and the text "Database Center for Life Science". A navigation menu includes "Wiki", "Timeline", "Roadmap", "Browse Source", "View Tickets", "New Ticket", "Search", and "Gantt Cha". The main heading is "BioHackathon 2008 meeting topics" with a sub-heading "New version". Below this is a list of topics.

BioHackathon 2008 meeting topics

New version

- [OpenBio* & Web Service Semantics](#) - Richard Bruskwiech
 - domain models & ontology: data types, service types and model API
- standardization
 - [Glycoinformatics](#) - Will York
 - [Interaction networks](#), DAS - Bruno Aranda
 - [Text Mining](#), Text mining - Kano
 - [Phyloinformatics](#) - Hilmar Lapp
- [Async services](#) - José M^a Fernández
 - WSRF?
- [Large data](#) - Tom Oinn

<http://hackathon.dbcls.jp/wiki/ListOfTopics>

- ・ オントロジー
 - ・ Open Bio Semantics 共通データ型 → ドメイン毎のモデルは相互運用性に必須、OpenBio* が UML で標準化？
- ・ 交換データの標準化
 - ・ Glycoinformatics (糖鎖) → 糖鎖データ解析フローの BioMoby/Taverna による実現
 - ・ 相互作用ネットワーク → PSI-MI を採用し IntAct など共通の WS を構築 Cytoscape で利用
 - ・ テキストマイニング → UIMA とバイオインフォの融合, iHOP は BioMoby に載っている
 - ・ Phyloinformatics (系統解析) → 入出力型の整理と WS のデザイン、NEXMLの利用促進
- ・ Async service (非同期通信) → BioMoby で採用されている WSRF は時代遅れ、WS-I に準拠しよう
 - ・ WSRF は Ruby で使えない、Soaplab にも非同期実装がある、WS-Addressing、WSO2 など
- ・ Large data (大規模データ) → SOAP アタッチメントは非現実的、リファレンスを渡して DL は bittorrent 等？
 - ・ リファレンスにデータ型の情報がつけられない、クライアントがデータを置く場所、データの生存期間の指定
- ・ セキュリティ → X509 電子署名によりユーザのデータ保護やサーバのデータ秘匿、計算機資源の割当などを管理
- ・ Open Bio* 開発グループ (BioPerl, BioRuby, BioJava) → BioSQL による共通のデータ交換プラットフォーム
- ・ REST or SOAP → WSDL 2.0 では統合可能
- ・ ワークフロークライアント → Taverna, Bio-jETI, Kepler, Remora, MOWServ, IWWE&M etc.
- ・ 相互運用性のあるワークフローの構築 → DDBJ/PDBj/KEGG を利用するワークフローの設計, DDBJ WABI
 - ・ Taverna は条件分岐が出来ない、連携するためのプログラミングが必要、WSの機能不足
- ・ BioMoby クライアント → G-language による Moby サービスのサポート、BioRuby によるクライアント

オントロジーと交換データ型の標準化

- ・ Open Bio Semantics 共通データ型
 - ・ バイオインフォの WS で交換される全データ型の標準化が望ましい
 - ・ 少なくともドメイン毎のデータモデルは必須
- ・ 個別ドメイン
 - ・ Glycoinformatics (糖鎖)
 - ・ 糖鎖データ解析フローの BioMoby/Taverna による実現
 - ・ 相互作用ネットワーク
 - ・ PSI-MI を採用し IntAct など共通の WS を構築、Cytoscape で利用
 - ・ Phyloinformatics (系統解析)
 - ・ NEXML の利用促進、入出力型の整理と対応する解析 WS のデザイン
 - ・ テキストマイニング
 - ・ UIMA とバイオインフォの融合、iHOP は BioMoby 上に載っている

交換データ型の標準化の流れ

- BioMoby のオブジェクトオントロジー
 - 似たようなオブジェクトが大量に登録されている (>500)
 - BLAST の出力など主要なデータ型はフラットに formatted-text 下
 - BioMoby サーバの構築が Perl/Java 以外では困難
- SOAP/WSDL (complexType) - XML schema
 - 標準データ型で定義し .xsd ファイルを共有 (open-bio.org でホスト?)
 - 同じオブジェクトモデルを Open Bio* とウェブサービス間で利用
- UML (BioUML?)
 - BioPerl, BioRuby, BioPython, BioJava で共通のモデルを設計
 - 相互運用性のあるオブジェクトモデルからコードを生成
 - もしくは Bio* は共通の C ライブラリに対するバインディングを使用

非同期通信 (Async service)


- ・ バイオインフォの解析フローでは時間のかかるステップも多い
 - ・ サーバにリクエスト→ジョブIDを発行→終わっているかポーリング
 - ・ 双方に WS サーバ機能があれば、WS-Addressing などで通知
- ・ BioMoby は WSRF を採用 (時代遅れになりつつある技術とも)
 - ・ WSRF は OGSA でも利用されている
 - ・ 実装が無く Ruby では利用できない
- ・ Soaplab にも非同期実装がある
- ・ → WS-I に準拠してほしい (WS-* などのスタックが必要)
 - ・ WSO2 (<http://wso2.org/>) の実装が良くなれば?

大規模データ (Large data) とセキュリティ

- ・ SOAP では MIME アタッチメントで送られがち
 - ・ Base64 エンコードされ一段と巨大化、全部届くまで解凍できない
- ・ データに対するリファレンスを渡して DL は REST や bittorrent 等で？
 - ・ リファレンスにはデータ型の情報がつけられない
 - ・ クライアントが解析のためのデータを一時的における場所の提供
 - ・ データの生存期間の指定
 - ・ 解析ワークフローではクライアントを介さずサーバ間で受け渡す
- ・ セキュリティ
 - ・ X509 電子署名
 - ・ ユーザのデータ保護、サーバのデータ秘匿、計算機資源 の割当などを管理

BioHackathon で残された課題

- ・ 現状の様々な問題点が提示された
 - ・ Standard data types and workflow
 - ・ Large data
 - ・ Async service
 - ・ Security
 - ・ Error handling
 - ・
- ・ Publication
 - ・ Manifesto 標準化の指標
 - ・ Journal article 成果のレビュー
- ・ 継続的な会議の開催ができれば
 - ・ BOSC session for web service
 - ・ annual BioHackathon

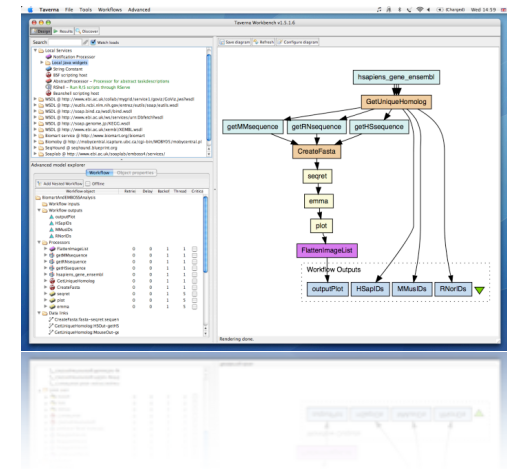


Open questions
Short term

- **Asynchronous services (& progress status notification)**
- **Error handling**
- **Replication (mirroring) system**
- **Daily Service quality control**
- **Service's help system**
- **Advanced service discovering**
- **Collections**
- **Large data sets (network overloading)**
- **Indirect (by reference) pass of data (locality of D & S)**
- **Workflows (storage in the repository, WWE-jmf)**
- **Semantic annotations (datatypes & Services)**
- **Define services categories**
- **New datatypes: Gene expression; images; ...**

相互運用性のあるワークフローの構築

- DDBJ/PDBj/KEGG を利用するワークフローの構築
 - Taverna を利用
 - 条件分岐が出来ない
 - 連携するためのプログラミングが必要
 - 現状の WS では解析のための機能が不足
- それぞれのサービスで何が可能か、何が不足しているか
 - 必要とされているワークフローの洗い出しと実装
 - 定期的なミーティングによる統合のための協力
 - 各サービスの品質向上
 - DDBJing, PDBjing, KEGGing などの講習会



統合にはほど遠く

- ・ データ型と解析ソフトウェアの数は増加し続けている
 - ・ それらに我々の標準化提案を強制することは現実的には困難
- ・ ウェブサービスにはスケールしない部分がある
 - ・ Large data - 生物学データの量は指数関数的に増加中
 - ・ Async service - 計算機資源はサーバに依存し限られている
 - ・ Workflow - プログラミングなしでパイプライン化することは困難
- ・ 反省と課題
 - ・ 個別技術要素の問題点は洗い出せたが、グローバルな結論はでていない
 - ・ 個々の技術についてはグリッドなどでもすでに議論されていた点が多い
 - ・ データやサービスの統一化には強力なイニシアチブと人員が必要
 - ・ 全体の分かる人が long term な目的を持って方向性を決めるべき
 - ・ 今のインフラは圧倒的に力不足 → サイエンスのできるウェブサービスへ

今後の統合にむけて何が可能か

- ・ 相互運用性の実現には継続的な国際コミュニティの努力が必要
 - ・ 主要なサービスプロバイダは相互運用性の高い API を提供すべき
 - ・ EBI, NCBI, DDBJ, KEGG, PDBj, CBRC, ...
 - ・ BioMoby
- ・ 非標準のサービスもクライアント側のソフトウェアで統合可能
 - ・ Open Bio* ライブラリ (BioPerl, BioRuby, BioJava, BioPython)
 - ・ Taverna, Seahawk, MOWServ, jORCA, ...
- ・ 日本では DBCLS において国内のサービスを統合する TogoWS を始動
 - ・ データベースやソフトウェア出力のパパーザやコンバータなどを提供
 - ・ BioMoby に対して WS-I 準拠のプロキシサーバを提供
 - ・ 国内外のサービスを統合したワークフローを集積、ドキュメント化

宣伝：TogoDB と TogoWS

- ・ DBCLS 基盤チームでのプロジェクト
 - ・ **TogoDB**: 受け入れのための deploy の非常に簡単なテンプレートDB
 - ・ ジェネリックな構成 - Ruby on Rails のプラグイン
 - ・ エクセルや CSV を 1 click で読み込み RDB を構築
 - ・ 統合/横断/絞り込み 検索のための GUI を自動生成
 - ・ 結果をじゃんじゃん見るための QuickLook 機能
 - ・ SOAP/WSDL による検索 API も同時に自動生成
 - ・ オープンソースで配布 / DBCLS でも受入サーバをホスト
 - ・ **TogoWS**: BioHackathon の成果をもとに国内の WS を中心に統合
 - ・ サイエンスの出来るインフラ整備につなげていく

LOGO

1. name table 2. upload csv 3. convert character code 4. create table 5. import data

table name
aaa

Upload file
C:\%a.txt
got 1 file (2 Bytes)

ukf
-w -x -d --no-best-fit-chars

Setup column information

enable	sample data	column name	label	type
<input checked="" type="checkbox"/>				string

create table

