BioHackathon 2008

for Web Service

DBCLS
Database Center for Life Science

Feb 11-15 @ Tokyo, Japan

http://ustream.tv/channel/biohackathon-2008-tokyo
http://ustream.tv/channel/biohackathon2008
http://hackathon.dbcls.jp/

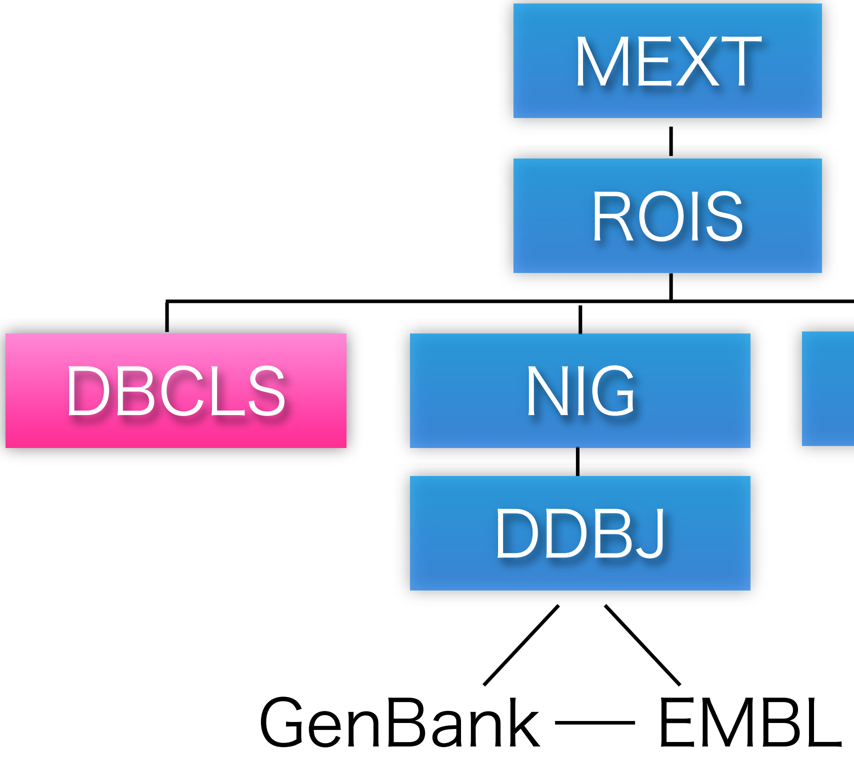
Toshiaki Katayama - organizer

- KEGG lab in Tokyo branch
  - KEGG API (SOAP/WSDL)
  - KEGG DAS
  - ortholog clusters, regulatory pathways, ...
- BioRuby project (since 2000)
- Japan Open Bio* Research Group (since 2004)
- DBCLS
  - Integrated databases
  - Integrated web services

DBCLS - sponsor

DBCLS
Database Center for Life Science

- Database Center for Life Science (since 2007)



MEXT

ROIS

DBCLS

NIG

NII

DDBJ

GenBank — EMBL

文部科学省 MINISTRY OF EDUCATION, CULTURE, SPORTS, SCIENCE AND TECHNOLOGY-JAPAN
http://www.mext.go.jp/

大学共同利用機関法人
情報・システム研究機構
Research Organization of Information and Systems
http://www.rois.ac.jp/

Research Organization of Information and Systems
National Institute of Genetics
http://www.nig.ac.jp/

Research Organization of Information and Systems
国立情報学研究所
National Institute of Informatics
http://www.nii.ac.jp/

:

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

BioHackathon #1-3

- 2002 BioHackathon #1, #2
  - Tucson - O'Reilly
  - South Africa - Electric Genetics

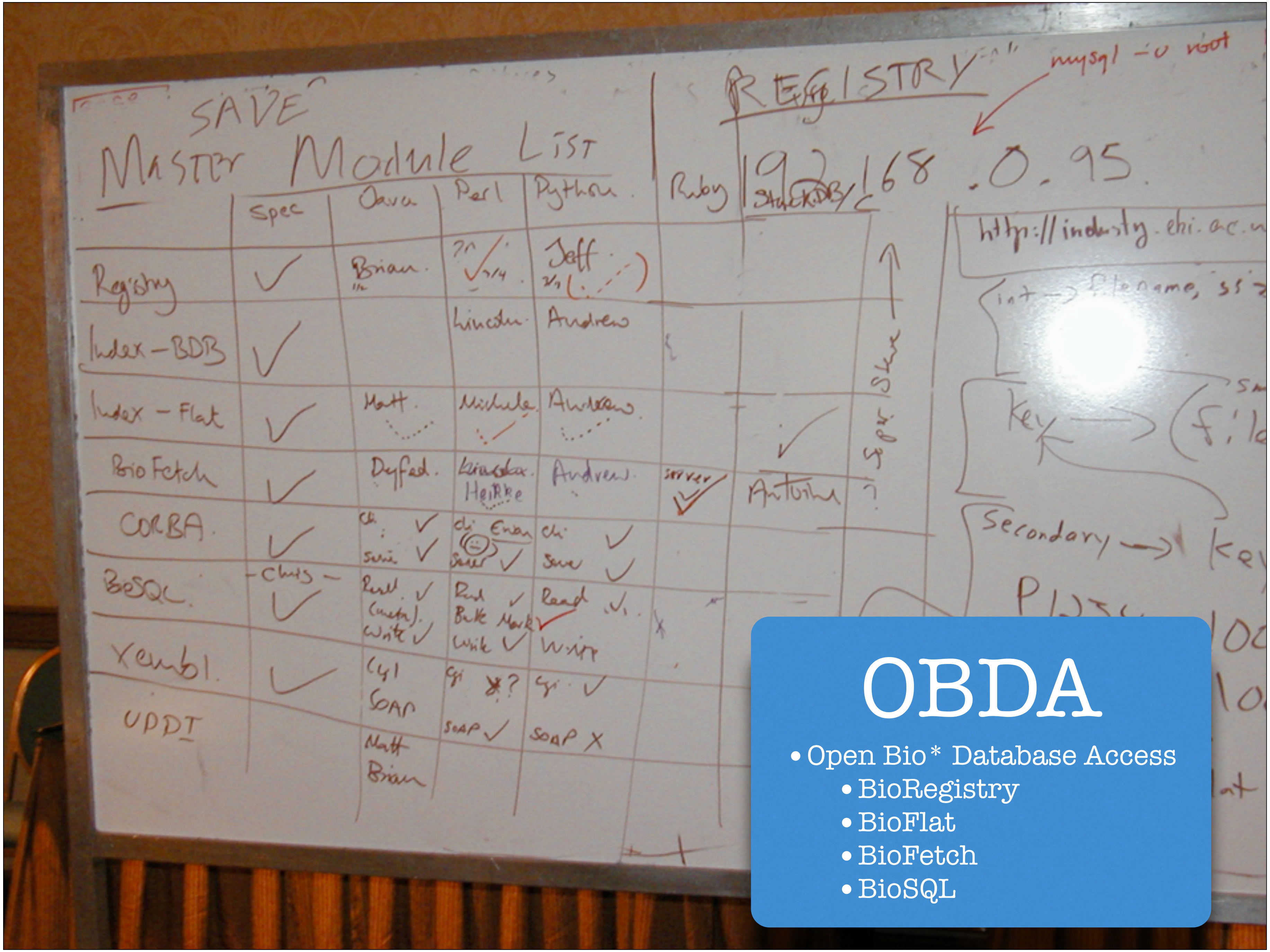
OBDA (Open Bio* Database Access)

- 2003 BioHackathon #3
  - Singapore - Apple

GBrowse (Bio::Graphics), Chado, BioMOBY, ...

BioHackathon #2

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



Master Module List

	Spec	Java	Perl	Python	Ruby	STACKDB/C
Registry	✓	Brian	✓	Jeff		
Index-BDB	✓		Lincoln	Andrew		
Index-Flat	✓	Matt	Michele	Andrew		✓
BioFetch	✓	Dyfed	[illegible] Heikke	Andrew	server ✓	Antonio
CORBA	✓	Chi ✓ Sean ✓	chi Ewan Sean ✓	chi ✓ Sean ✓		
BioSQL	-Chris- ✓	Read ✓ (meta). Write ✓	Read ✓ Bulk Mark Write ✓	Read ✓ Write	X	
Xembl	✓	Cgi SOAP	Cgi ? SOAP ✓	Cgi ✓ SOAP X		
UDDI		Matt Brian				

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 Bruskiewich
- 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

Participants

Service provider

Open Bio*

Client developer

BioMOBY

Exchange data format standardization

Favor to ask

- Discussion
  - please speak slowly & clearly
- Summary
  - please log to Wiki
- Be creative
  - find your own goals by yourself :-)

BioHackaton 2008 Open Space topics

- Ontology - Richard Bruskiewich
    - data types, service types
- standardization
    - Glycoinformatics - Will York
    - Interaction networks, DAS - Bruno Aranda
    - Text mining - Kano
    - Phyloinformatics - Hilmar Lapp
- Async service - Jose, Fernandez
    - WSRF?
- Large data - Tom Oinn
    - SOAP attachements
- Security - Oswaldo Trelles
- Description & Discovery - Mark Wilkinson
- Open Bio*
    - BioSQL - Mark Schreiber

- Writing (creating) interoperable workflows - Tom Oinn, Stuart Owen
- Writing the manifesto

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

BioHackathon2008
Meeting report - DBCLS/BioRuby

Toshiaki Katayama

Human Genome Center, Univ of Tokyo
ktym@hgc.jp

Primary intention

■integration {名-1} : 統合、一体化、統一、融合、調和、集大成

- Mission of the "Database Center for Life Science"
  - Integration of the domestic life science databases

Database Center for Life Science

- My proposals
  - Generate a template database which is really easy to deploy
    - Plugin for "Ruby on Rails"
      - Load .xls or .csv file by 1 click and build RDB
      - Automatically generate web interface to search
      - Automatically generate SOAP/WSDL APIs also
    - We can also host the databases at DBCLS
  - Integrate large systems such as KEGG, DDBJ, PDBj via web service
    - Review the best way to achieve this during the BioHackathon

Togo
DB

Togo
WS

Togo WS - standard objects to exchange?

- BioMoby object ontology
  - too many similar objects
  - not so easy to build BioMoby server in different language like Ruby (only usable to Perl and Java clients)

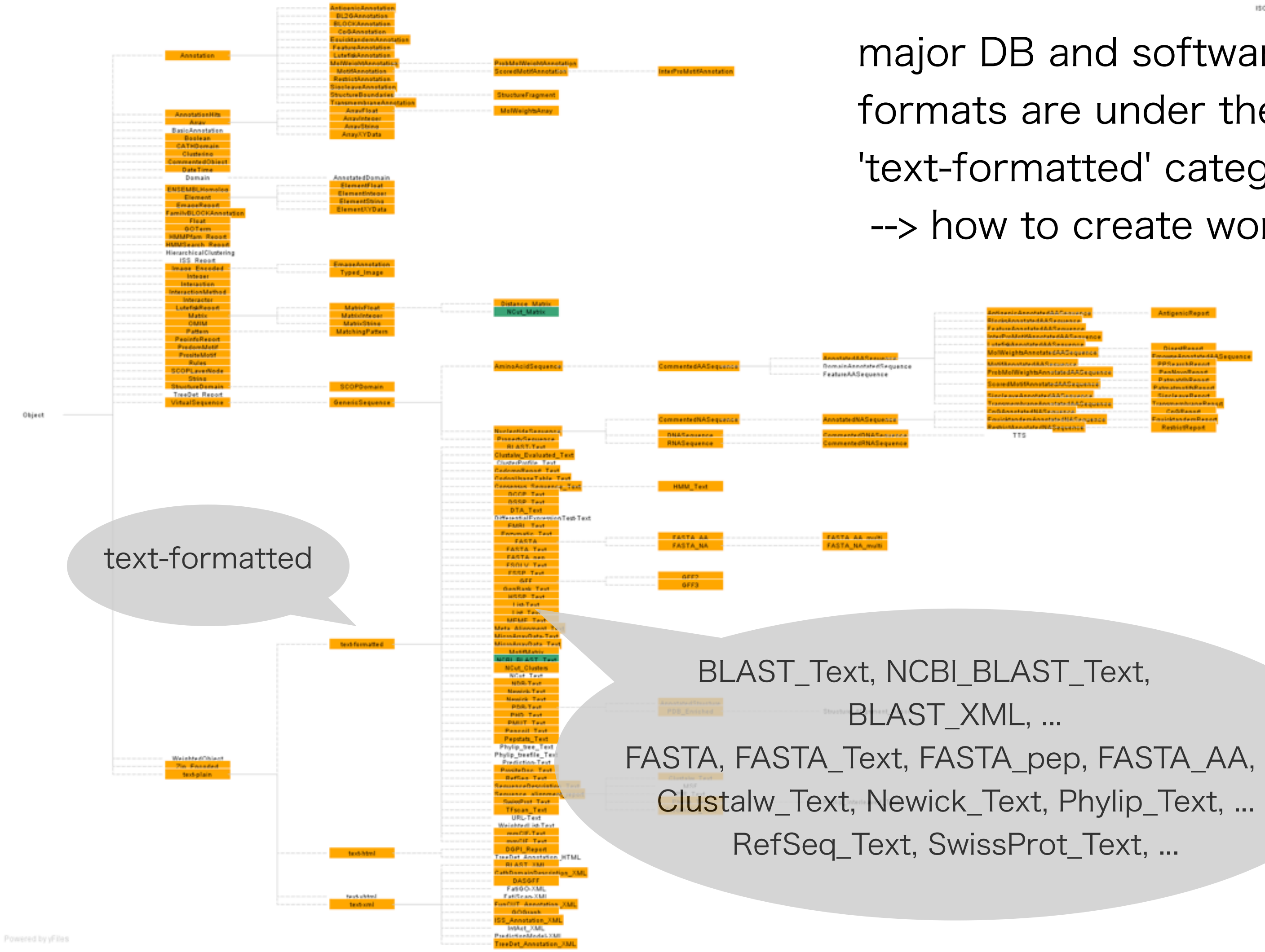
- SOAP/WSDL (complexType) - XML schema
  - put standard .xsd files at open-bio.org
  - share the same object models among Bio* projects and WSs

- UML (BioUML?)
  - truly interoperable object models among Bio* projects
  - written in C
  - bindings in each Bio* library

BioMoby/MOWServ object ontology

major DB and software output formats are under the single 'text-formatted' category flatly!

--> how to create workflows?



text-formatted

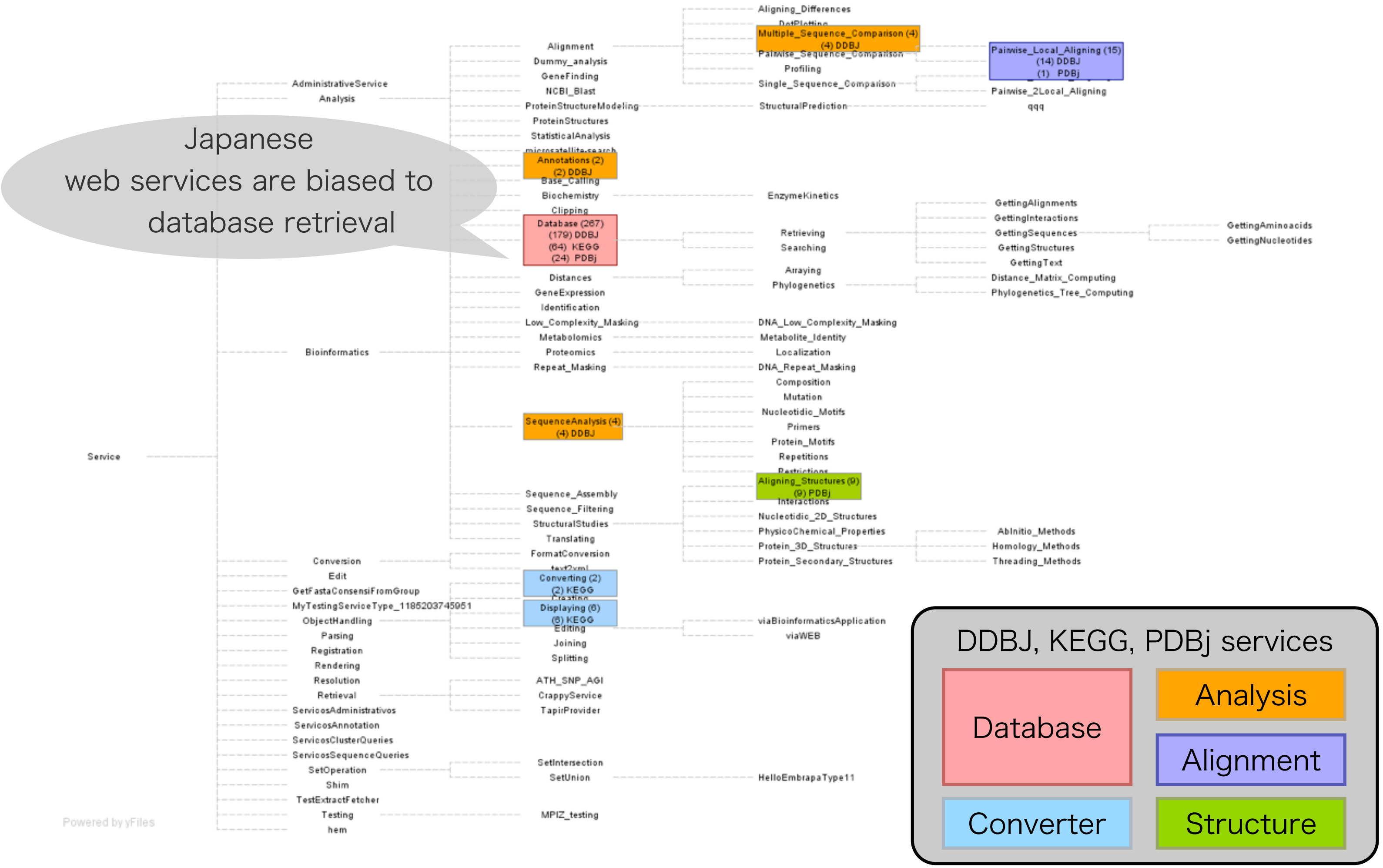
BLAST_Text, NCBI_BLAST_Text, BLAST_XML, ...

FASTA, FASTA_Text, FASTA_pep, FASTA_AA, ...

Clustalw_Text, Newick_Text, Phylip_Text, ...

RefSeq_Text, SwissProt_Text, ...

BioMoby service ontology



Japanese
web services are biased to
database retrieval

DDBJ, KEGG, PDBj services

🟥 Database
🟧 Analysis
🟦 Alignment
🟦 Converter
🟩 Structure

Far from integration

- The number of programs and data types keep growing
  - Force them to comply our standardization effort is impossible
- Web service won't scale
  - Large data - Amount of data is increasing exponentially
  - Async service - Computational power is also limited
  - Workflow - Pipelining services is still hard w/o programming

Future integration

- Continuous community effort is required for the interoperability

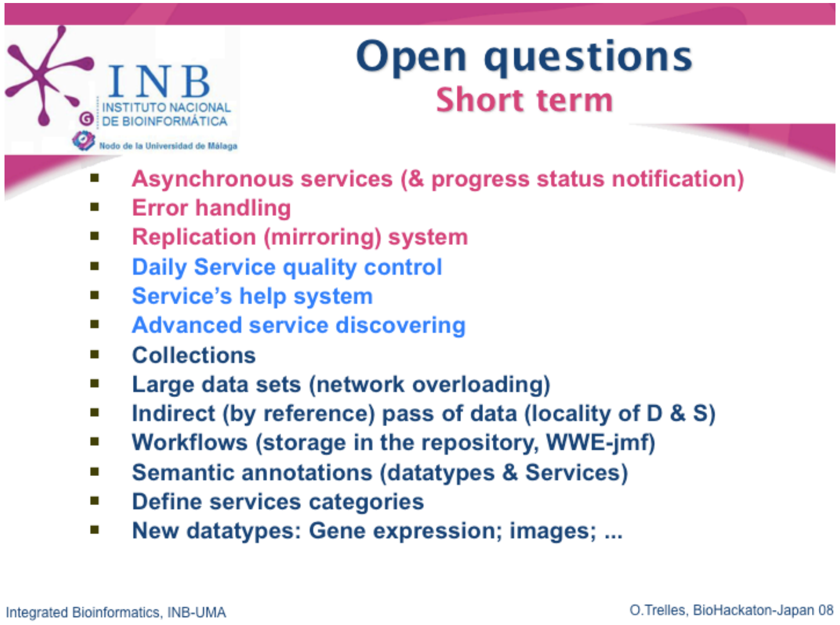
- Large scale services should provide interoperable APIs
  - EBI, NCBI, DDBJ, KEGG, PDBj, CBRC, ...
  - BioMoby

- Non-standard services can also be integrated by clients and libraries
  - Open Bio* libraries (BioPerl, BioRuby, BioJava, BioPython)
  - Taverna, Seahawk, MOWServ, jORCA, ...

- In Japan, DBCLS will develop TogoWS to integrate domestic services
  - Provide "shim" services: parsers and converters for various db/app
  - Provide WS-I proxy for the BioMoby services?

Remaining issues

- Number of problems are represented
  - Standard data types and workflow
  - Large data
  - Async service
  - Security
  - Error handling
  - .....
- Publication
  - Manifesto
  - Journal article
- Meeting
  - BOSC session for web service
  - annual BioHackathon



INB
INSTITUTO NACIONAL
DE BIOINFORMÁTICA
Nodo de la Universidad de Málaga

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; ...

Integrated Bioinformatics, INB-UMA

O.Trelles, BioHackaton-Japan 08

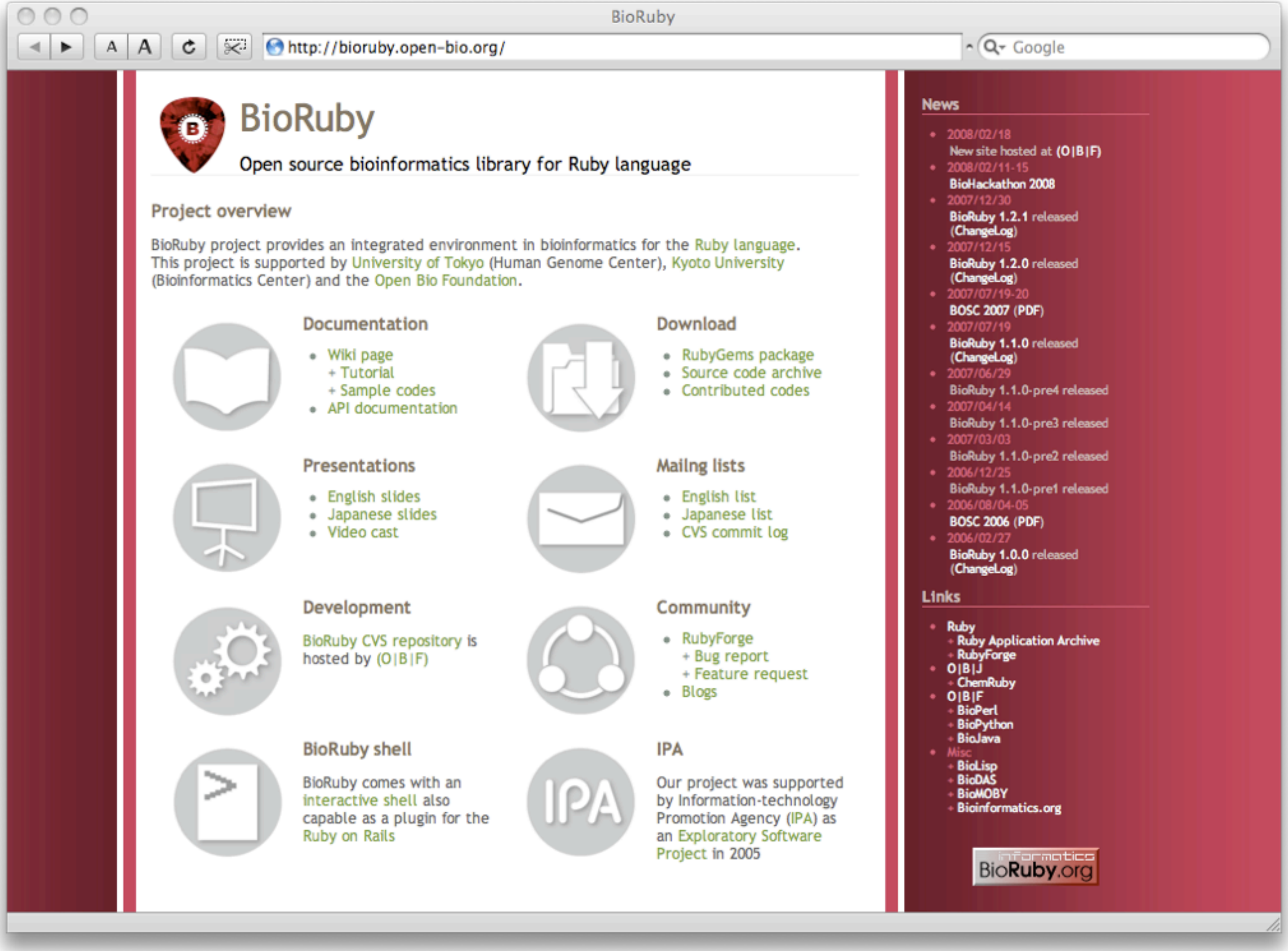
Slide by Prof. Trelles (INB/MOWServ)

BioRuby group - hackathon results

- BioMoby client is newly developed
  - access to BioMoby central to find services and execute
- Bio::Sequence class is extended to be generic
  - inport/export interface for major sequence data formats
- Bio::SQL is enhanced and tested consistency
  - based on ActiveRecord (Ruby on Rails)
- Some bug fixes

- Web site migration
  - open-bio.org - Nice new top page + Media Wiki
  - rubyforge.org - gem distribution, bug trackers

- Details -> http://hackathon.dbcls.jp/wiki/BioRuby



BioRuby

Open source bioinformatics library for Ruby language

Project overview

BioRuby project provides an integrated environment in bioinformatics for the Ruby language. This project is supported by University of Tokyo (Human Genome Center), Kyoto University (Bioinformatics Center) and the Open Bio Foundation.

Documentation

- Wiki page
  - Tutorial
  - Sample codes
- API documentation

Download

- RubyGems package
- Source code archive
- Contributed codes

Presentations

- English slides
- Japanese slides
- Video cast

Mailng lists

- English list
- Japanese list
- CVS commit log

Development

BioRuby CVS repository is hosted by (O|B|F)

Community

- RubyForge
  - Bug report
  - Feature request
- Blogs

BioRuby shell

BioRuby comes with an interactive shell also capable as a plugin for the Ruby on Rails

IPA

Our project was supported by Information-technology Promotion Agency (IPA) as an Exploratory Software Project in 2005

News

- 2008/02/18
  New site hosted at **(O|B|F)**
- 2008/02/11-15
  BioHackathon 2008
- 2007/12/30
  BioRuby 1.2.1 released
  (**ChangeLog**)
- 2007/12/15
  BioRuby 1.2.0 released
  (**ChangeLog**)
- 2007/07/19-20
  BOSC 2007 (**PDF**)
- 2007/07/19
  BioRuby 1.1.0 released
  (**ChangeLog**)
- 2007/06/29
  BioRuby 1.1.0-pre4 released
- 2007/04/14
  BioRuby 1.1.0-pre3 released
- 2007/03/03
  BioRuby 1.1.0-pre2 released
- 2006/12/25
  BioRuby 1.1.0-pre1 released
- 2006/08/04-05
  BOSC 2006 (**PDF**)
- 2006/02/27
  BioRuby 1.0.0 released
  (**ChangeLog**)

Links

- Ruby
  - Ruby Application Archive
  - RubyForge
- O|B|J
  - ChemRuby
- O|B|F
  - BioPerl
  - BioPython
  - BioJava
- Misc
  - BioLisp
  - BioDAS
  - BioMOBY
  - Bioinformatics.org

informatics BioRuby.org