BioHackathon for interpreting biological knowledge with Semantic Web technologies

Opening Remarks

Database Center for Life Science

Prof. Takagi
Director of DBCLS

BioHackathon
8-12 Feb. 2010
DBCLS / AIST Tokyo Japan

BioHackathon 2010 Symposium

>80 participants!

32 invited

33 domestic

15 organizers

Hackathon Basics

BioHackathon
= bio + hack + marathon
= bioinformatics coding camp!

- One unique style of the international workshop
  - Face-to-face meeting of the geeks
  - Addressing issues by intensive software development
  - Originated from free software culture
- History
  - Open Bio* Hackathon (2002, 2003)
  - Phyloinformatics Hackathon (2006-)
  - Systems Biology Super Hackathon (2008)
  - DBCLS Hackathon (2008-)

Evolution of DBCLS BioHackathons

Mission of the DBCLS
= integration of life science resources
(standardization / interoperability)

The 1st DBCLS BioHackathon (2008)
Towards interoperable web services in life science with Open Bio * libraries

The 2nd DBCLS BioHackathon (2009)
Integration of web services in bioinformatics applications

The 3rd DBCLS BioHackathon (2010)
Interpretation of biological knowledge with Semantic Web technologies

Goals of the BioHackathon 2010

Learn the Semantic Web technology
- *What is the RDF, OWL, SPARQL, ... #@&*%$?!*
- *Is it promising?*
- *Will it give better answer than existing systems?*

Survey the current status and develop some software
- *Linked Data*
  - What kind of biological data is currently available?
- *Triple Stores*
  - Which implementation to use? pros/cons? Scalability?
- *Semantic Web Tools in Open Bio**
  - Generic tools to manipulate RDF graph
  - Interface to acess SPARQL endpoints

BH10 schedule: Day 1

Session	Time	Presented by / Participants
Talk	10:10	Erick Antezana
Talk	10:40	Matthias Samwald
Talk	11:00	Thomas Kappler and Jerven Bolleman
Talk	11:20	François Belleau
Talk	11:40	Heiko Horn
Lunch	12:00	
Talk	13:30	Tetsuro Toyoda
Talk	13:50	Mark Wilkinson
Talk	14:10	Andrea Splendiani
Talk	14:30	Mitsuteru Nakao and Toshiaki Katayama
Poster	15:00	Invited participants +α
Open space	16:30	All participants
Banquet	18:00	Registered participants

BH10 schedule: Day 2-5

Day	Session	Time	Participants / Activities
2/9	Hack	9:00 - 18:00	Invited & domestic participants
2/10	Hack	9:00 - 18:00	Invited & domestic participants
2/11	Hack	9:00 - 18:00	Invited & domestic participants
2/12	Hack	9:00 - 16:00	Invited & domestic participants
2/12	Talk	16:00 - 18:00	Summary session
2/12	Drink	18:00 - 21:00	Farewell party :-)
2/13-15	Write	10:00 - 18:00	Writing a meeting report with volunteers

10:10-10:40

Belgium 

Erick Antezana

AFFILIATION

Genomics Data Platform
BioInformation Management
Bayer CropScience

PROJECTS

Semantic Systems Biology
http://www.semantic-systems-biology.org
Cell Cycle Ontology
http://www.cellcycleontology.org

Towards a Semantic Systems Biology:
Biological Knowledge Management Using
Semantic Web Technologies

10:45-11:05

Austria 

Matthias Samwald

AFFILIATION

DERI Galway (Ireland) and
the Konrad Lorenz Institute (Austria)

PROJECTS

Concept Web Alliance
http://conceptweblog.wordpress.com/

<a>Tag
http://hcls.deri.org/atag/generator/

High-level knowledge representation on the Semantic Web: the Concept Web Alliance and related efforts

11:05-11:25

Switzerland

Thomas Kappler / Jerven Bolleman

AFFILIATION

UniProt in RDF

UniProt ConsortiumUniProt Consortium
Swiss Institute of Bioinformatics
http://uniprot.org

11:25-11:45

Canada 

François Belleau

AFFILIATION

Laval University

PROJECTS

Bio2RDF
http://www.bio2rdf.org
Semantic web atlas of postgenomic knowledge

Bio2RDF cognoscope : A killer app for the life science

11:50-12:10

Denmark 

Heiko Horn

AFFILIATION

NNF Center for Protein Research
Faculty of Health Sciences
University of Copenhagen

PROJECTS

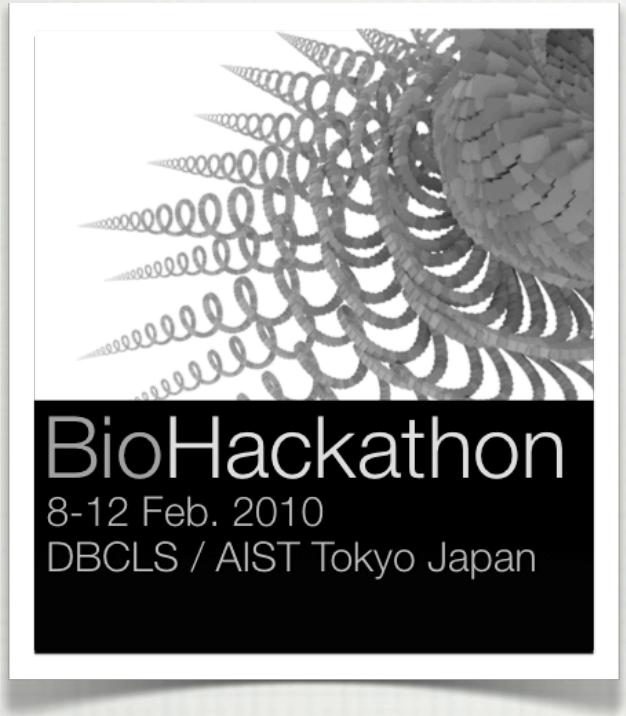
Reflect
http://www.reflect.ws
NetworKIN
http://networkin.info

Reflect - text mining in Semantic Web

13:35-14:05

Japan 

Tetsuro Toyoda



AFFILIATION

Bioinformatics and Systems Engineering division,
RIKEN

PROJECTS

SciNeS
http://omicspace.riken.jp/base/index.html#6

RIKEN SciNeS

14:05-14:25

Canada 

Mark Wilkinson

AFFILIATION

Dept. of Medical Genetics,
University of British Columbia,
Heart + Lung Institute at
St. Paul's Hospital

PROJECTS

BioMoby
http://www.biomoby.org
SADI
http://sadiframework.org
Cardio SHARE
http://biordf.net/cardioSHARE

SADI - semantic web service

14:25-14:45

UK 

Andrea Splendiani

AFFILIATION

Centre for Mathematical and
Computational Biology
Rothamsted Research

PROJECTS

RDFScape
http://bioinformatics.org/rdfscape
Ondex
http://www.ondex.org
BioPAX
http://www.biopax.org
Beat
http://beat.sourceforge.net
SWAT4LS
http://www.swat4ls.org

Visualization and analysis of biological networks on the Semantic Web

Extending TogoWS and Open Bio* libraries for Linked Data

Mitsuteru Nakao (DBCLS)
Toshiaki Katayama (HGC, Univ. of Tokyo)

Personal Semantic Web
--- If you have data, export it as Linked Data ---

http://hackathon3.dbcls.jp/

BH10 Preliminary Objectives

Queries

Targeted biological queries to be resolved by the Semantic Web technologies.

Datasets

List currently available and/or still missing datasets (as Linked Data / RDF) to resolve the queries.

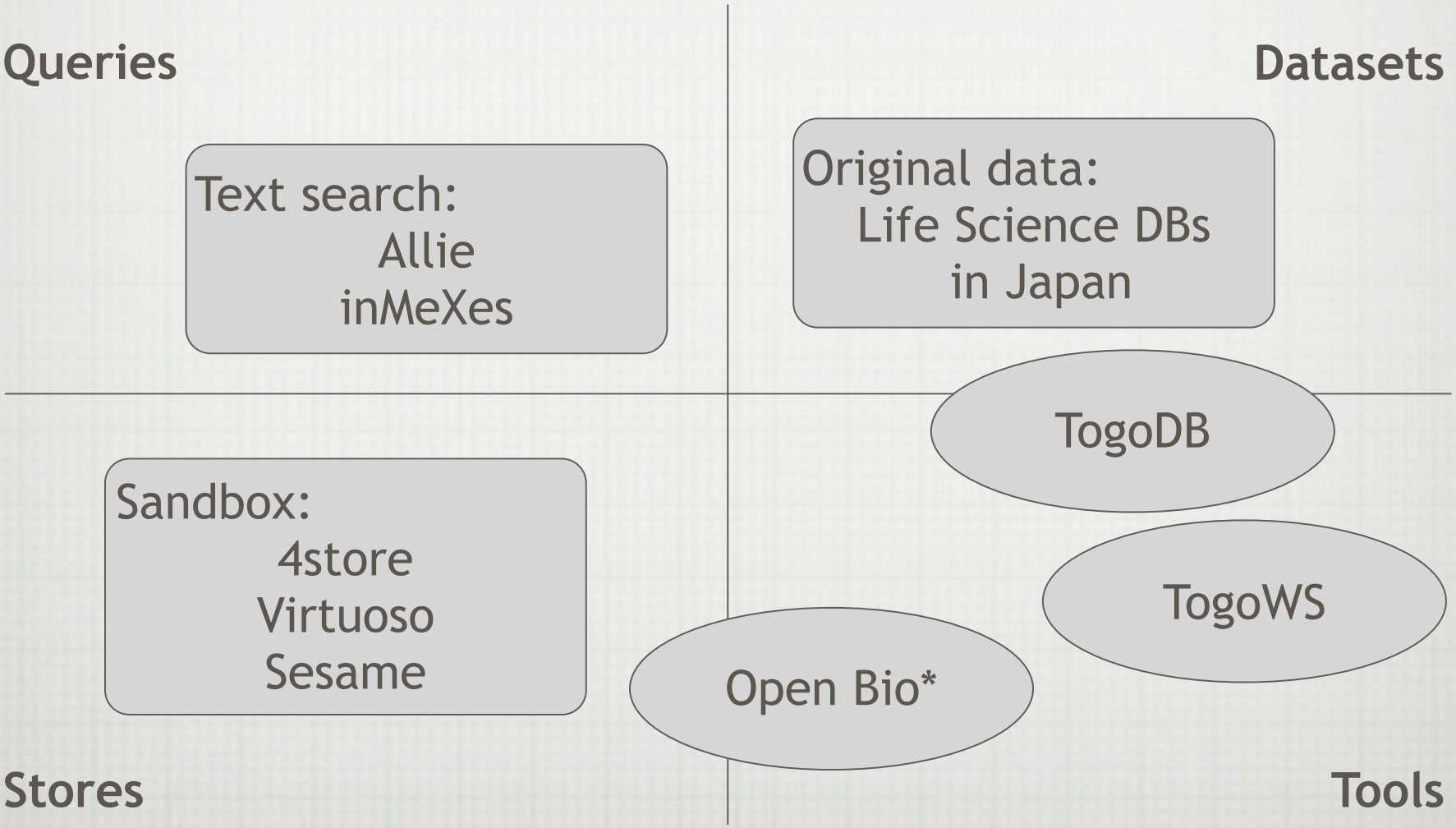
Develop our own or survey existing extensible storage systems for RDF triples and functional query interfaces.

Stores

Develop common APIs among Open Bio* projects for RDF stores and SPARQL endpoints.

Tools

Current efforts in DBCLS

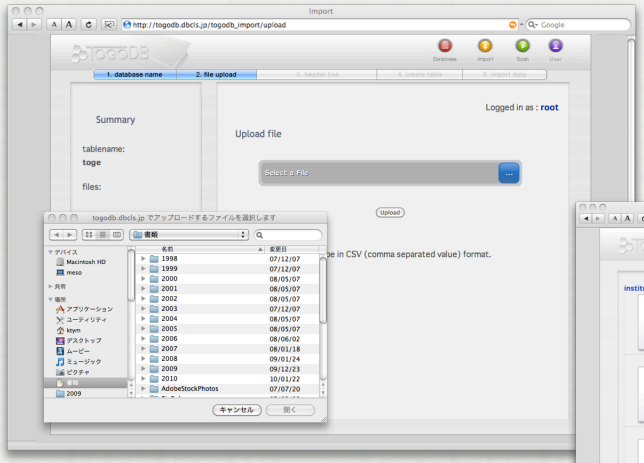


SemWeb for everybody

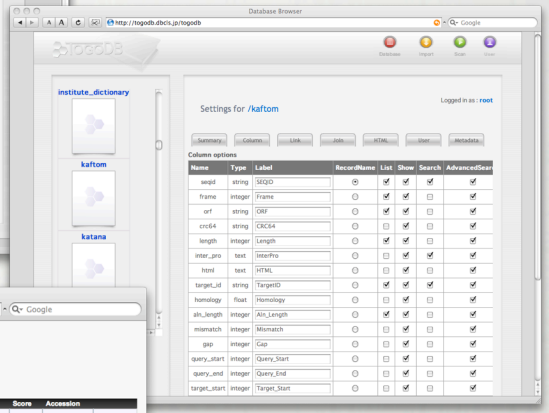
TogoDB

deploy your DB in 5min

upload

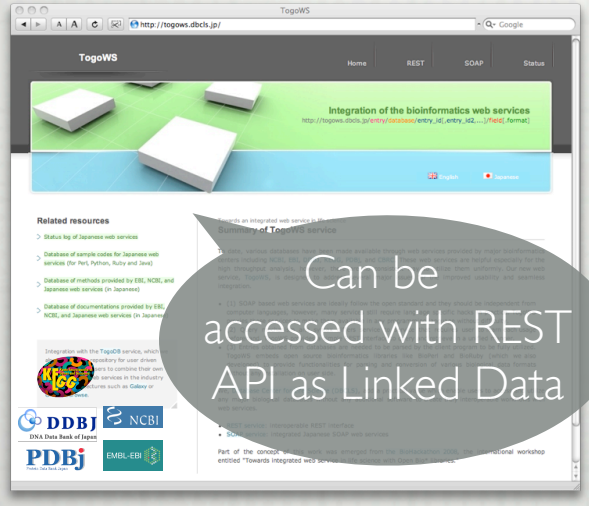


configure



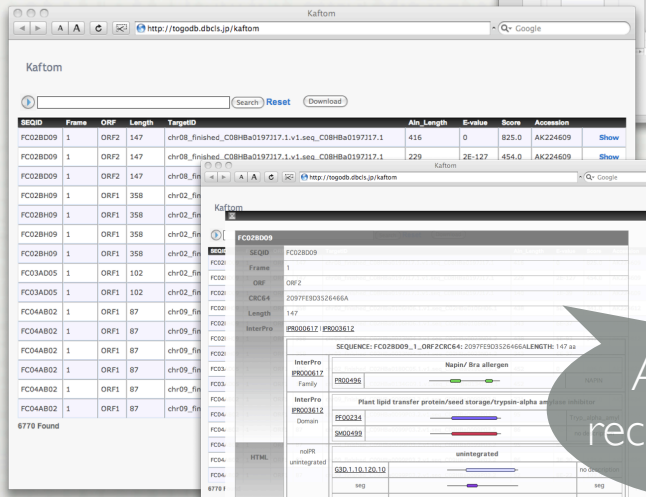
TogoWS

integrated REST/SOAP services



endpoint

Can be accessed with REST API as Linked Data



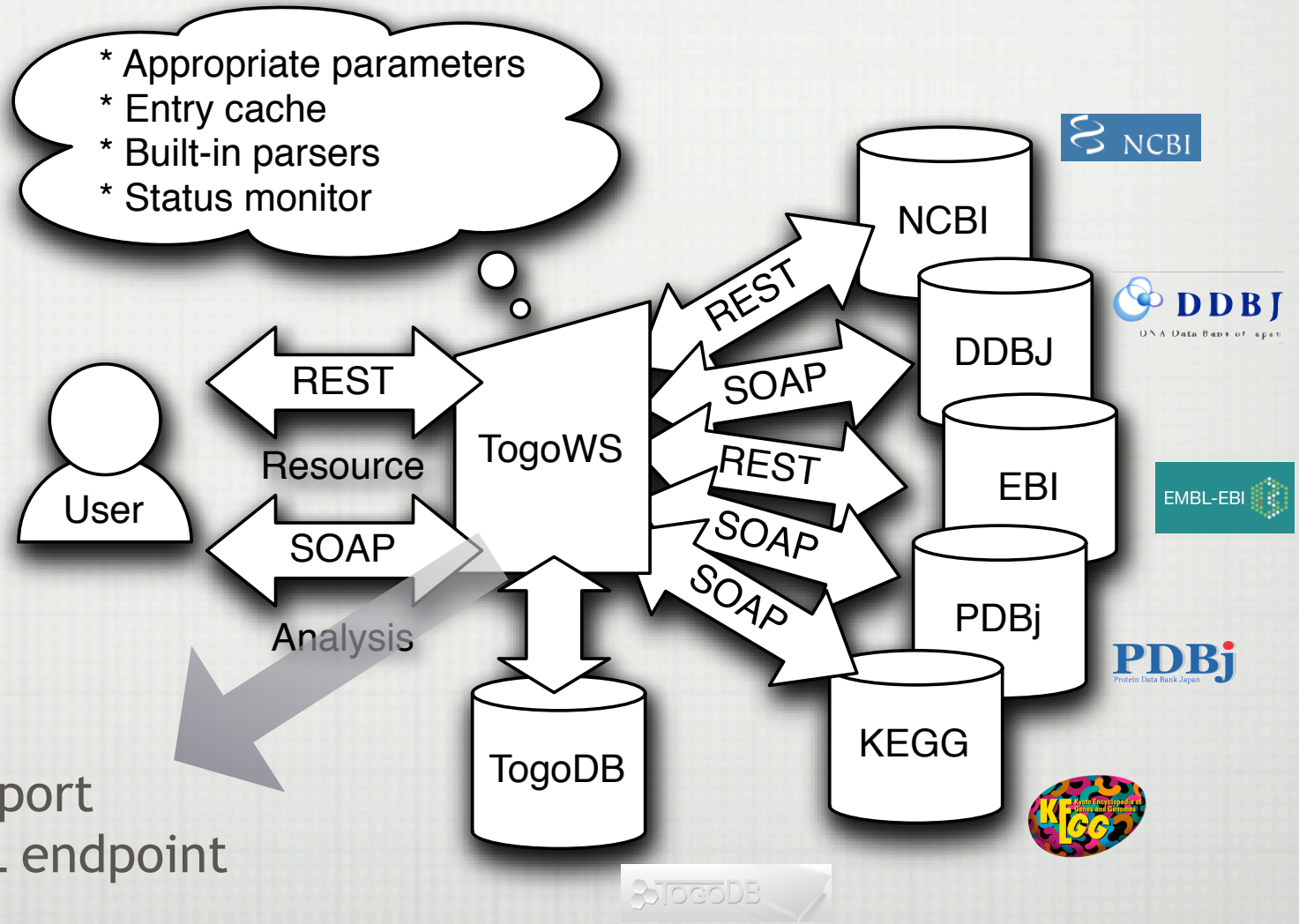
deploy

Any database record by the user

http://togows.dbcls.jp/entry/togodb-foobardb/123.rdf

http://togodb.dbcls.jp/foobardb/show/123

As a RDF/SPARQL provider



TODO:
RDF export
SPARQL endpoint

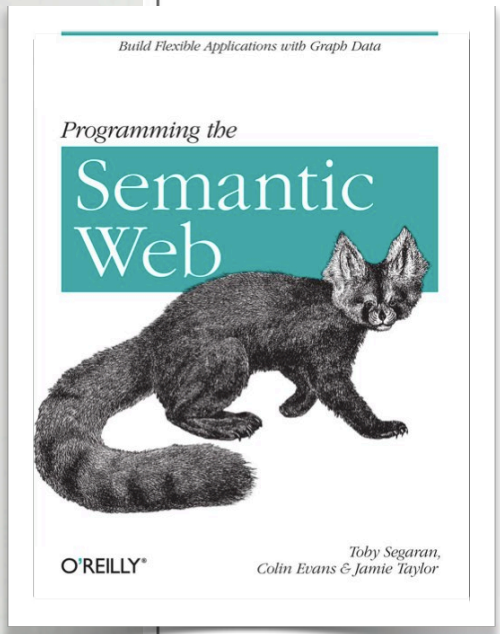


Figure 11-1. W3C semantic web technology stack

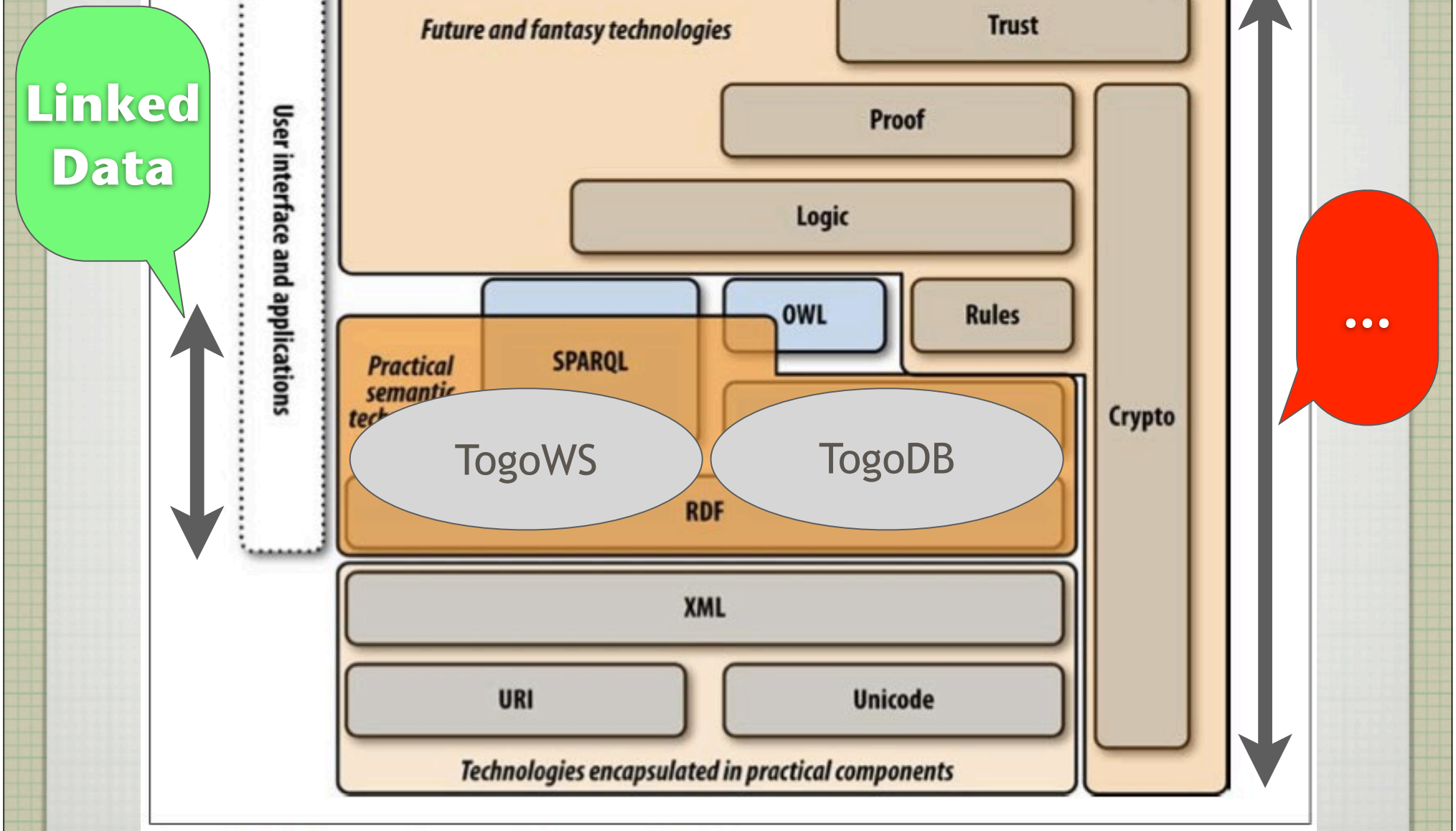


Figure 11-2. A practical view of the semantic stack

BRIEFINGS IN BIOINFORMATICS. VOL 10. NO 4. 392–407
Advance Access publication May 19, 2009

doi:10.1093/bib/bbp024

Biological knowledge management: the emerging role of the Semantic Web technologies

Erick Antezana, Martin Kuiper and Vladimir Mironov

Submitted: 5th January 2009; Received (in revised form): 17th April 2009

Table 1: Projects using Semantic Web technologies within the life science domain

Project	Keywords	Technologies	Web site	Reference
Bio2RDF	Mashup, linked data, global warehouse, complex queries	RDF, SPARQL	http://bio2rdf.org/	[165]
BioDash	Disease, compounds, therapeutic model, pathway	RDF, OWL	http://www.w3.org/2005/04/swls/BioDash/Demo/	[166]
BioGateway	Semantic systems biology, hypothesis generation	RDF, SPARQL	http://www.semantic-systems-biology.org/biogateway/	[167]
CardioSHARE	Collaborative, distributed knowledgebase, reasoning, web services	RDF, SPARQL	http://cardioshare.icapture.ubc.ca/	[120]
Cell-Cycle Ontology (CCO)	Cell cycle, protein–protein interactions, reasoning, ontology patterns	RDF, OWL, SPARQL	http://www.cellcycleontology.org/	[168]
CViT	Cancer, tumor, gene-protein interaction networks	RDF	https://www.cvit.org/	[135]
FungalWeb	Fungal species, enzyme substrates, enzyme modifications, enzyme retail	OWL		[169]
GenoQuery	Genomic warehouse, mixed query, tuberculosis	RDF, SPARQL	http://www.lri.fr/~lemoine/GenoQuery/	[170]
HCLS W3C	Knowledge base, life sciences, prototype	RDF, OWL, SPARQL	http://www.w3.org/TR/hcls-kb/	[171]
Kno.e.sis	Nicotine dependence, biological pathway	RDF, SPARQL, OWL	http://knoesis.wright.edu/research/semsci/application_domain/sem.life.sci/bio/research/	[172]
Linked Life Data	Pathways, interactions	OWL	http://www.linkedlifedata.com	[173]
LinkHub	Document ranking, text categorization, query corpus	RDF	http://hub.gersteinlab.org/	[174]
Lipid bibliosphere	Lipids, metabolites, reasoning	OWL		[128]
Neurocommons	Uniform access, package-based distribution	RDF, SPARQL	http://neurocommons.org/	[175]
RDFScape	Systems biology, cytoscape, reasoning	RDF, SPARQL	http://www.bioinformatics.org/rdfscape	[132]
S3DB	Lung cancer, omics	RDF	http://www.s3db.org/	[176]
SWAN - AlzPharm	Neuromedicine, alzheimer, neurodegenerative disorders	RDF, OWL	http://swan.mindinformatics.org	[177, 178]
SEMMAS	Web services, intelligent agents	OWL	http://semmas.inf.um.es/prototypes/bioinformatics.html	[179]
SOMWeb	Distributed medical communities	RDF, OWL	http://www.cs.chalmers.se/proj/medview/somweb/	[129]
Thea-online	Protein interactions, annotations, pathways	RDF, SPARQL	http://bioinfo.unice.fr:8080/thea-online/	[180]
yOWL	Yeast, phenotypes, interactions	OWL	http://ontology.dumontierlab.com/yowl-hcls	[181]

friendfeed

Search

Advanced search
Tools

The Life Scientists: Attila Csordas

Bioinformaticians: which is the bioinformatics project (on the web) which embodies best 'the semantic web' as we think of it today?

July 8, 2008 - Comment - Share

You, Nakao M., Konrad Förstner and 3 other people liked this (Un-like)

bio2rdf http://www.bio2rdf.org/ ? - Pierre Lindenbaum

UniProt is one of the first life sciences databases to distribute all of their data in RDF format (both via FTP and the Web, ~1B triples) -- that ought to count for something :-) - Eric Jain

Semantic web (as defined by the W3C) != speaking toasters. It's "just" data (RDF/OWL) that is accessible on the Web. Bonus: Provide a SPARQL endpoint so people can query the data rather than just retrieve it by URI. Challenges: Data modeling, scalability (if you have a lot of data), and creating generic-yet-usable end user tools that work directly with the RDF graph data model (I have yet to see any of these)... - Eric Jain

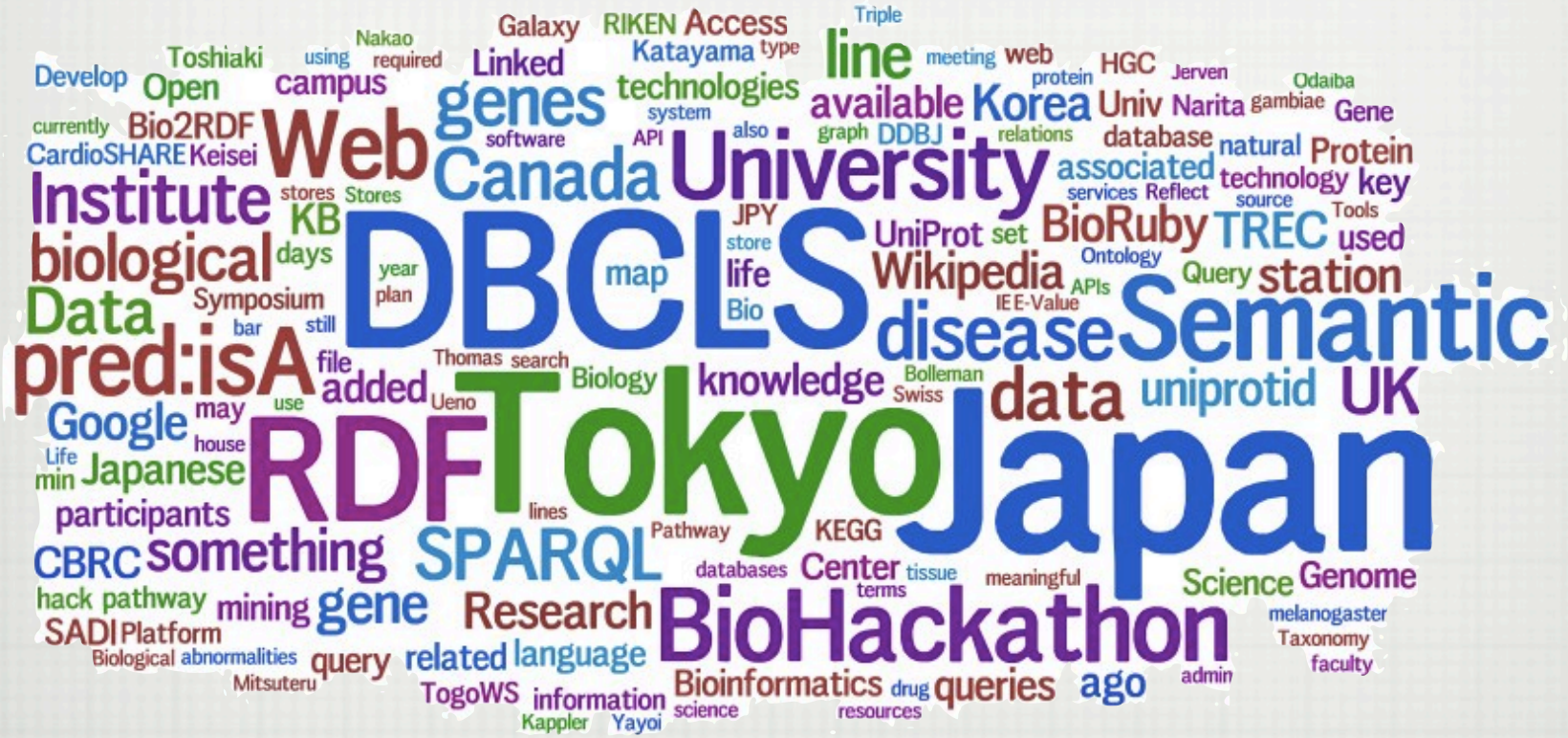
RDF, OWL, OBO, SPARQL endpoints WHATEVER, doesn't really matter. One of the most important things is agreeing on, defining and sharing vocabularies. Easy to say, much harder

Semantic web (as defined by the W3C) != speaking toasters. It's "just" data (RDF/OWL) that is accessible on the Web. Bonus: Provide a SPARQL endpoint so people can query the data rather than just retrieve it by URI. Challenges: Data modeling, scalability (if you have a lot of data), and creating generic-yet-usable end user tools that work directly with the RDF graph data model (I have yet to see any of these)... - Eric Jain

Open Bio* ?

http://friendfeed.com/the-life-scientists/188942c1/bioinformaticians-which-is-bioinformatics

What's come out?

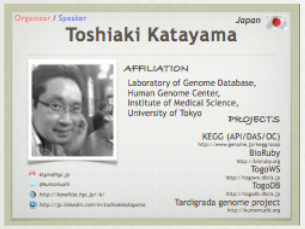


What kind of achievements will emerge from the cloud?

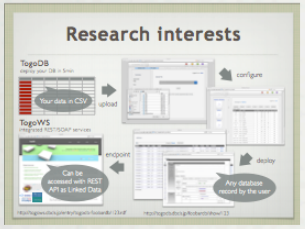
Nobody knows. Let's look again after one week. :-)

Poster session 15:00-

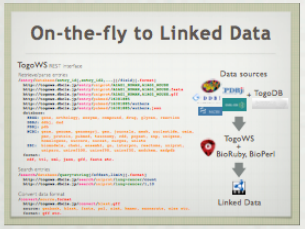
Poster presentation



← Organizers will pin up the cover sheet on the wall



← Please hang your slides under the cover sheet



Self-introduction
Look around freely
Talk to each other
Find your collaborators for BH10

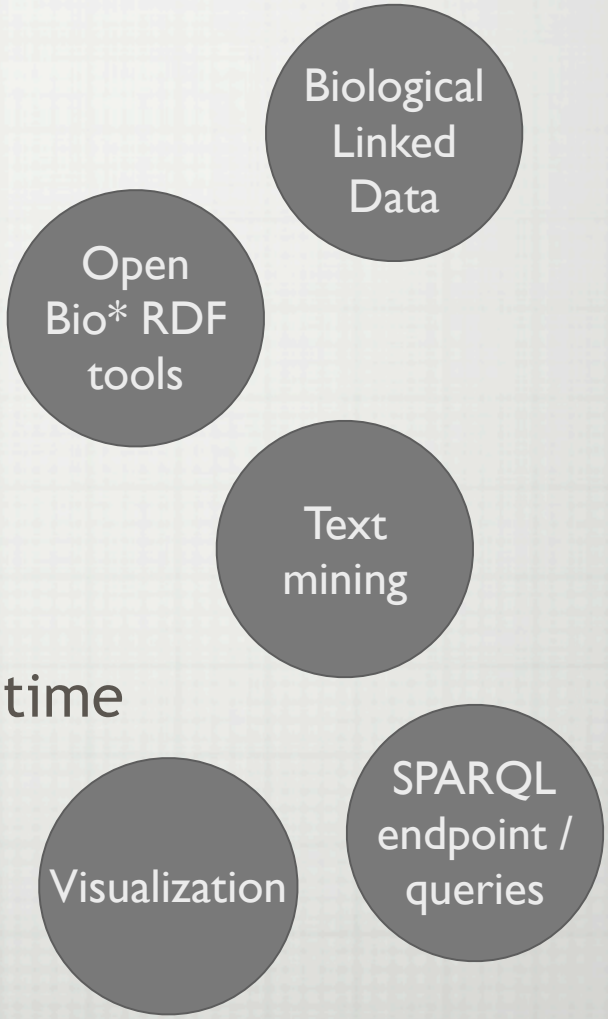
If you forget to bring your printed poster, ask organizers nearby. We have printers.

Open space session 16:30-

Open space discussion
= Sort of brainstorming
in self-organized discussion groups

Procedure / Rules

1. Raise candidate discussion topics
2. Select some discussion groups
3. Start discussion in group
4. Look around some discussions
5. You can switch group at any time
6. You can create additional group at any time
7. Finally join at least one group
8. Sort out issues
9. What will you challange during BH10?



Open space session 16:30-

16:50-17:20

UseCases!

Arakawa

Network / interaction

Kawaji

Visualization

Andrea

NextGen Seq

Alberto

Text mining

17:25-17:55

Implementation bootcamp

Rutger

Francois

Biological Linked Data

SPARQL endpoint / queries

How to properly use OWL

Jerven

Katayama

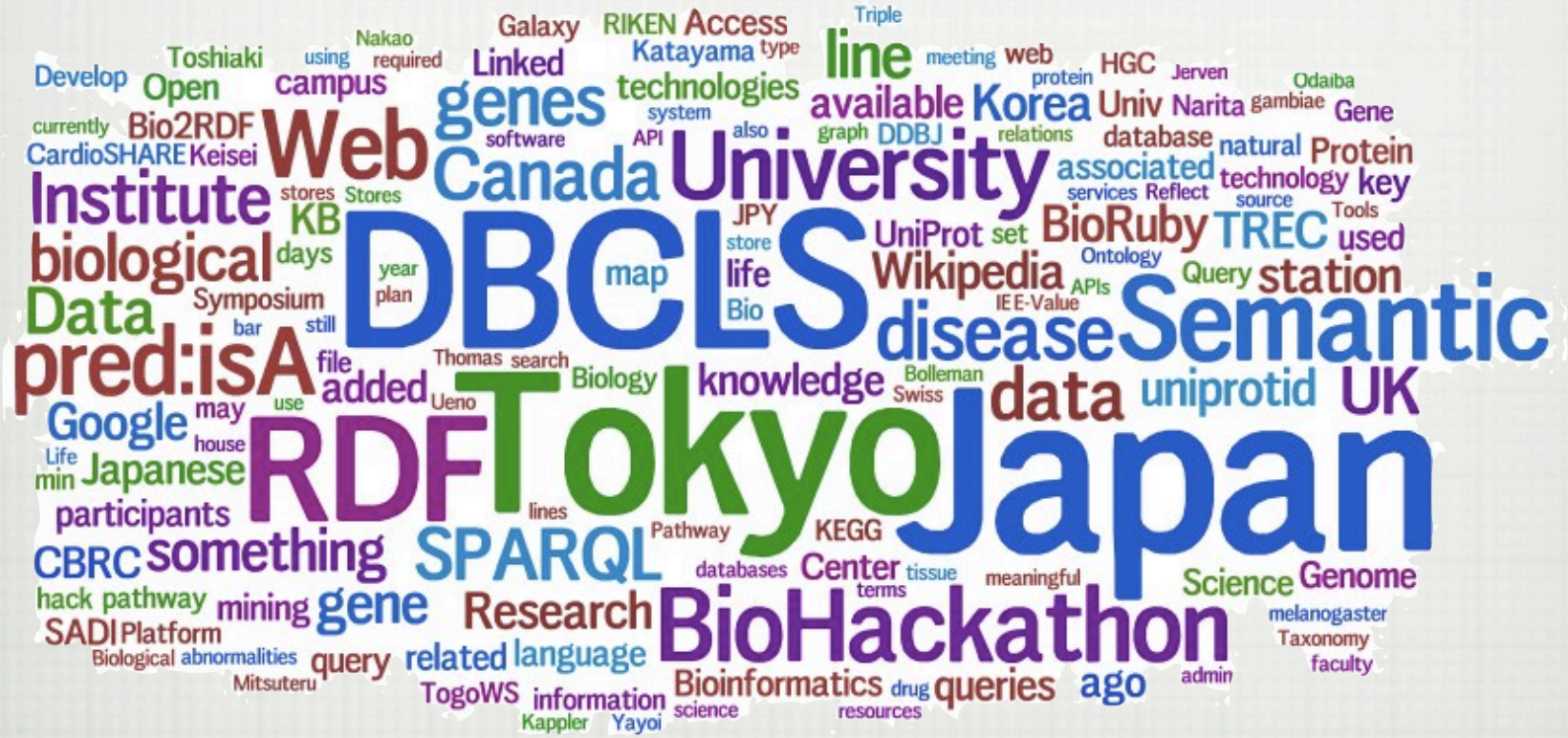
Open Bio* RDF tools Ruby/Python

DB provider: versioning LSID?

BioHackathon for interpreting biological knowledge
with Semantic Web technologies

BH2010
Summary
Session
12 Feb 16:00-

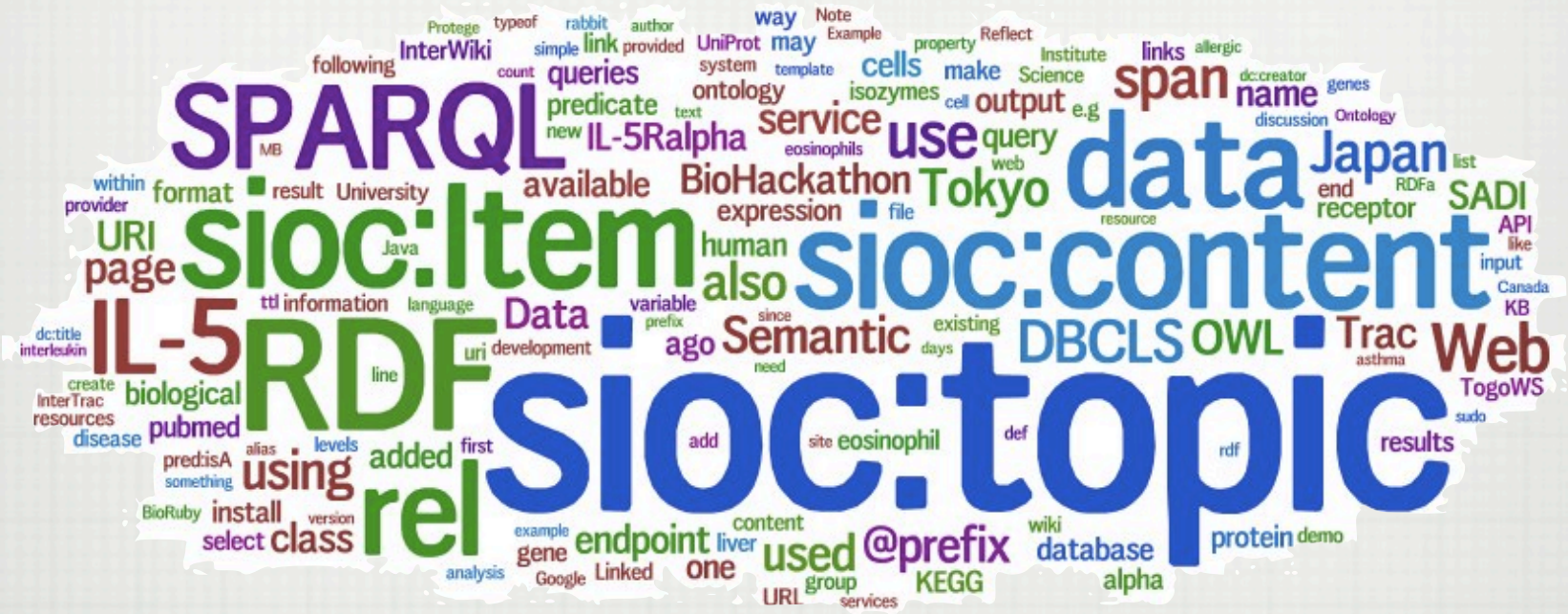
Before BH10



What kind of achievements will emerge from the cloud?

Nobody knows. Let's look again after one week. :-)

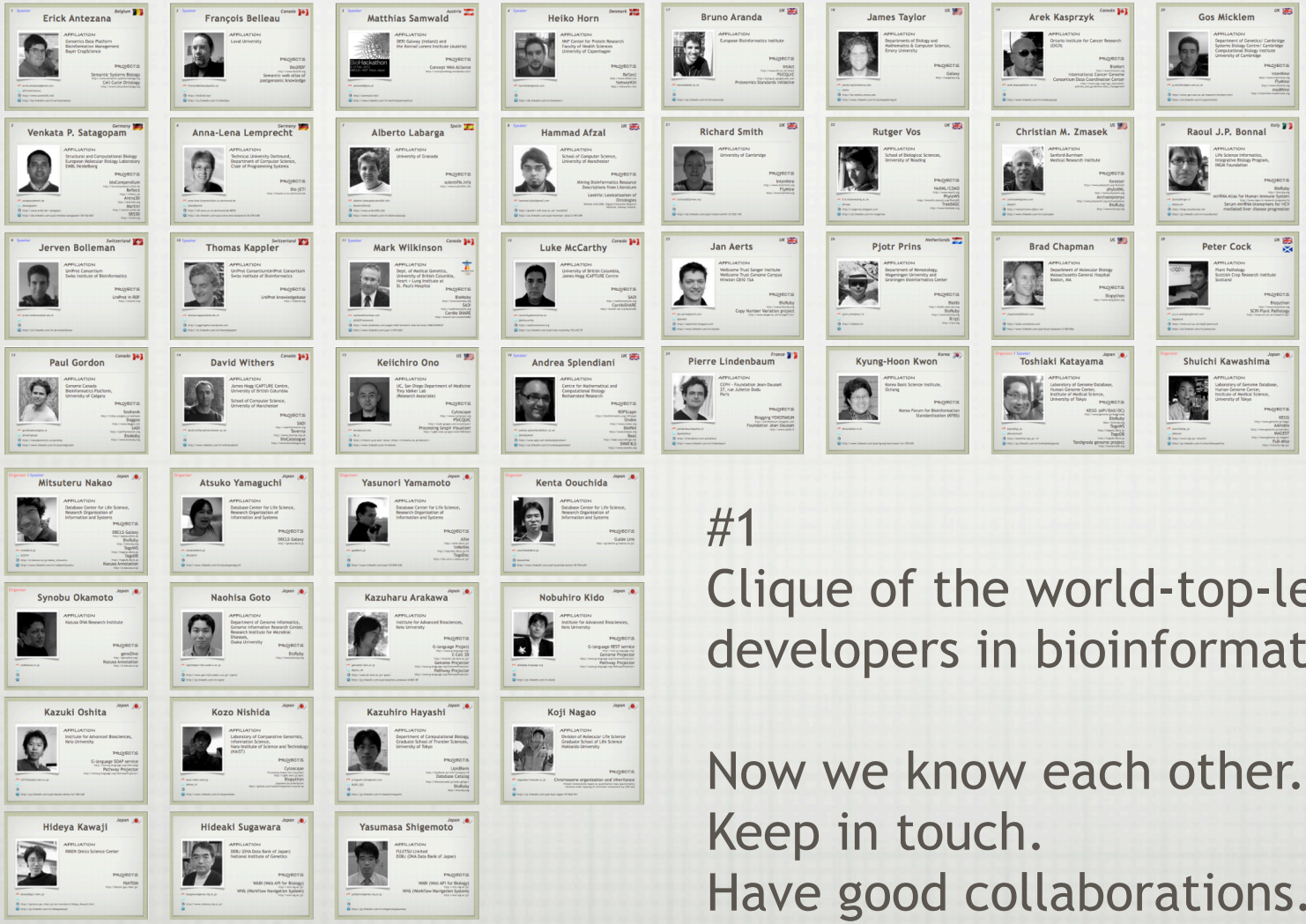
After BH10



Now we have brain-washed with RDF and SPARQL! :-)

Geek++

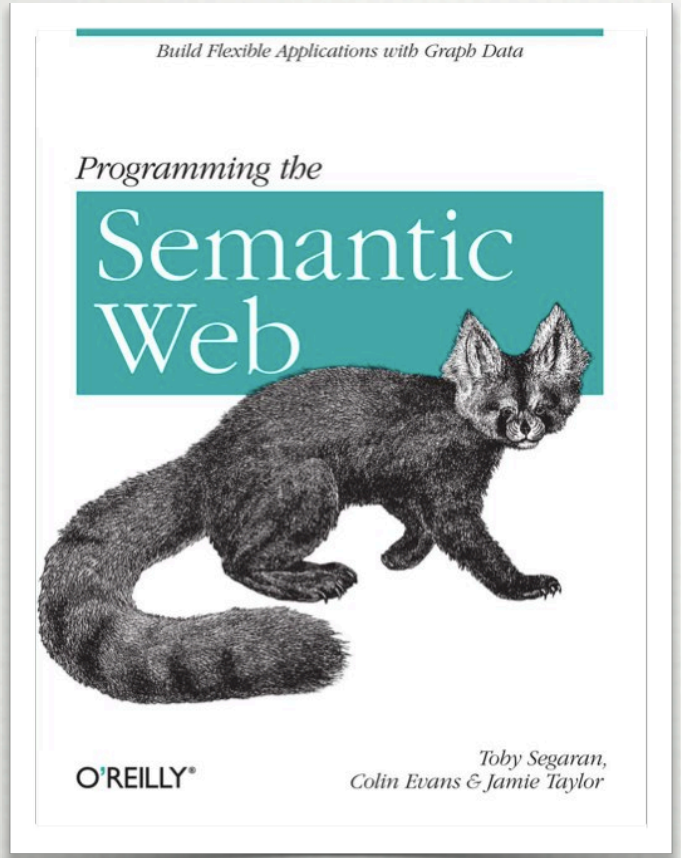
What we made/learned actually?



#1
Clique of the world-top-level developers in bioinformatics.

Now we know each other.
Keep in touch.
Have good collaborations.

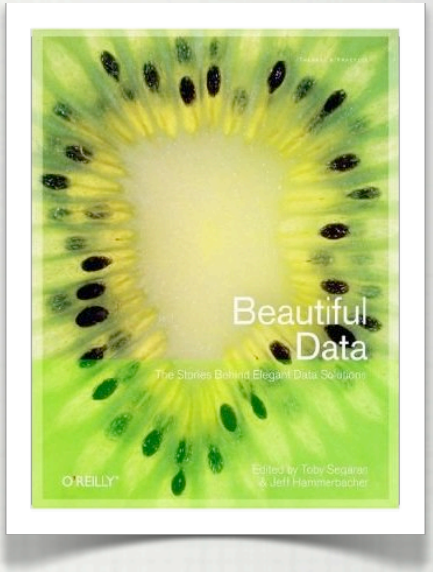
What we made/learned actually?



Name space

Tools

Hosting



#2
Database providers now understand potential of SemWeb (RDF/SPARQL) technologies

What we made/learned actually?

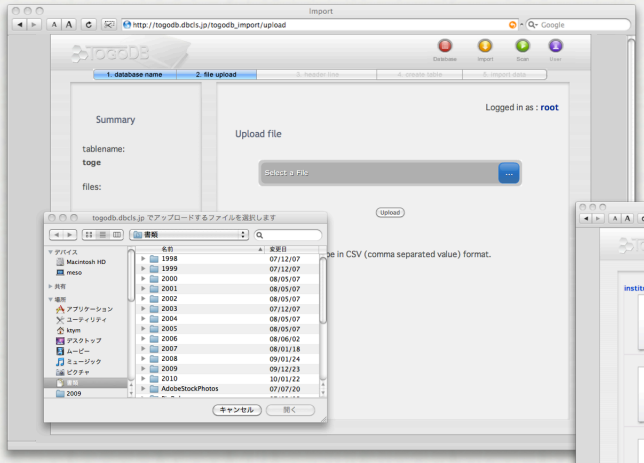
#3
Different tastes

SemWeb for everybody

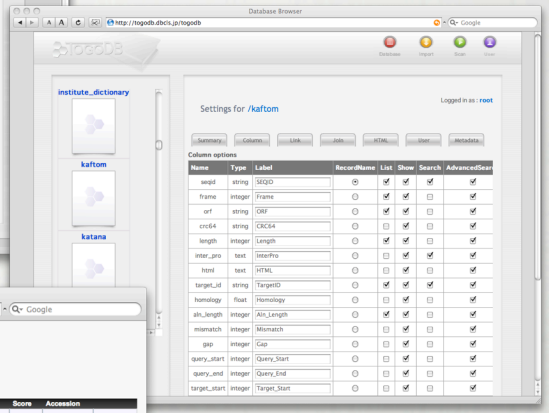
TogoDB

deploy your DB in 5min

upload

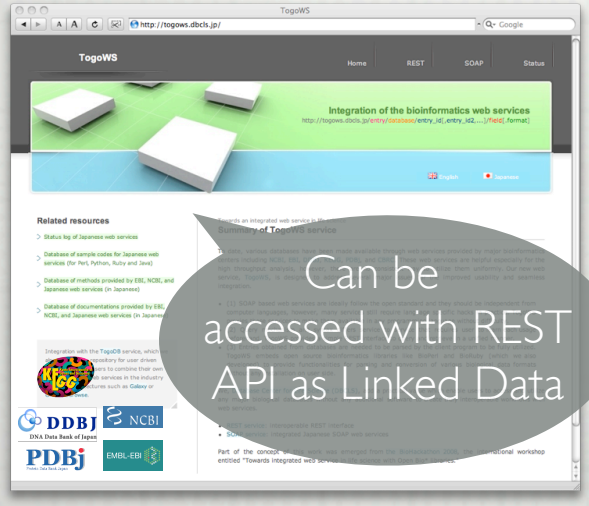


configure



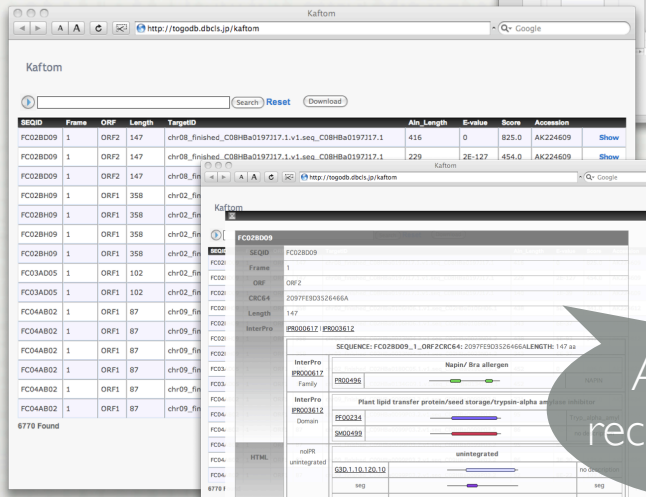
TogoWS

integrated REST/SOAP services



endpoint

Can be accessed with REST API as Linked Data



deploy

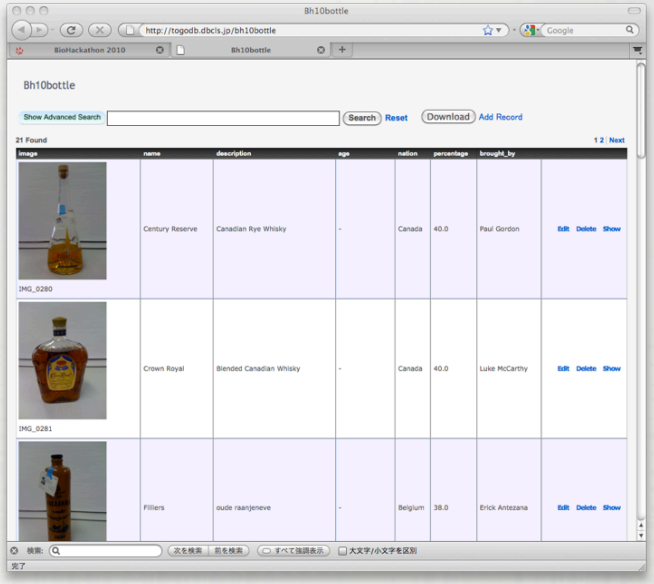
Any database record by the user

http://togows.dbcls.jp/entry/togodb-foobardb/123.rdf

http://togodb.dbcls.jp/foobardb/show/123

DEMO

Expose user's data to
the Semantic Web world

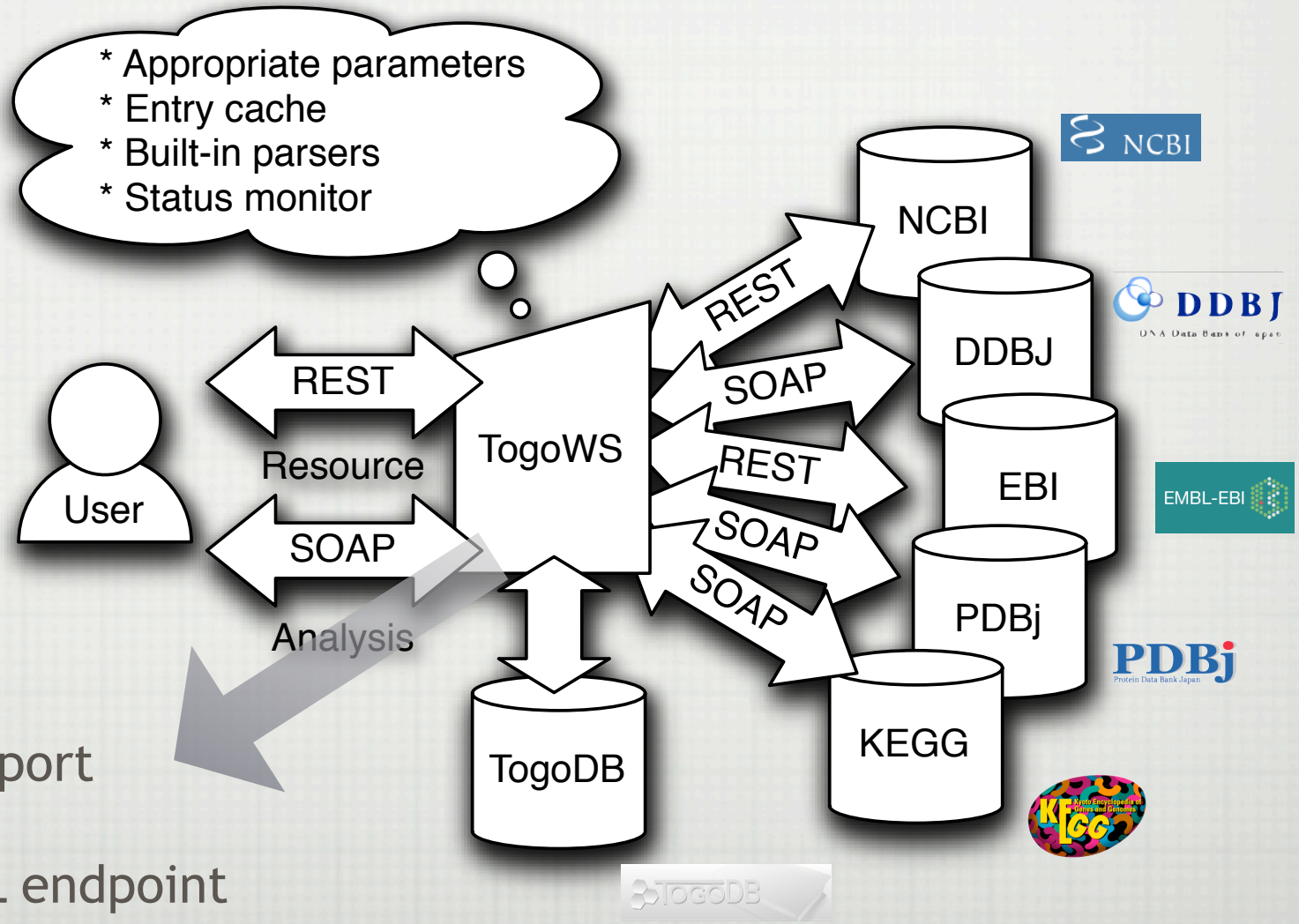


http://togodb.dbcls.jp/dbname/show/1

http://togows.dbcls.jp/entry/togodb-dbname/1.xml

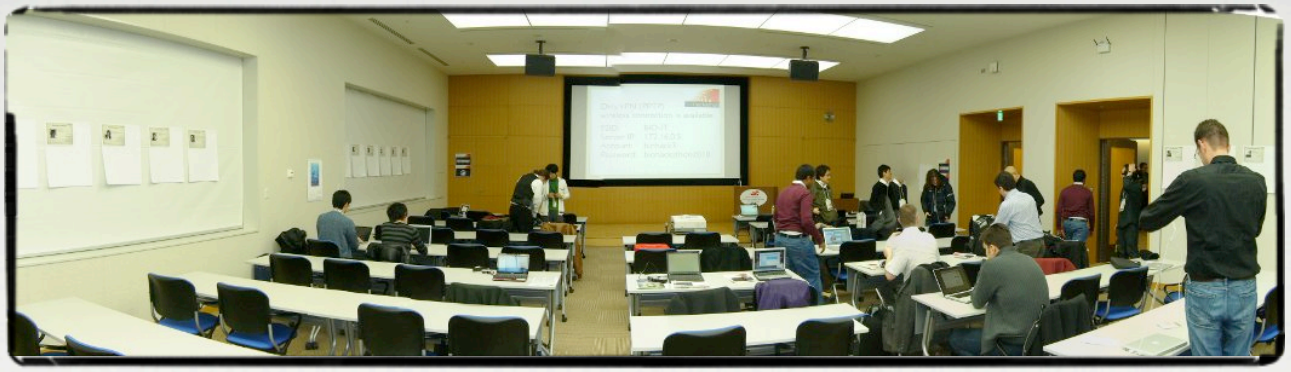
http://togows.dbcls.jp/entry/togodb-dbname/1.ttl

As a RDF/SPARQL provider



DONE:
RDF export
TODO:
SPARQL endpoint

BIOHACKATHON 2010 GROUP PHOTO



TSUKIJI FISH MARKET TOUR

ASAKUSA TOUR

Special Thanks to

- Keiko Sakuma, Eiji Ueda, Toshihisa Takagi and DBCLS staff
- Keiko Nemoto, Kiyoshi Asai and CBRC staff
- Alan Ruttenberg (NeuroCommons) through ML discussions
- Organizers
  - Mitsuteru Nakao, Shuichi Kawashima,

    Shinobu Okamoto, Shin Kawano

    Yasunori Yamamoto, Atsuko Yamaguchi