Moby, Legacy Apps and the Semantic Web

Web Services BioHackathon 2008

Paul Gordon
Sun Center of Excellence for Visual Genomics
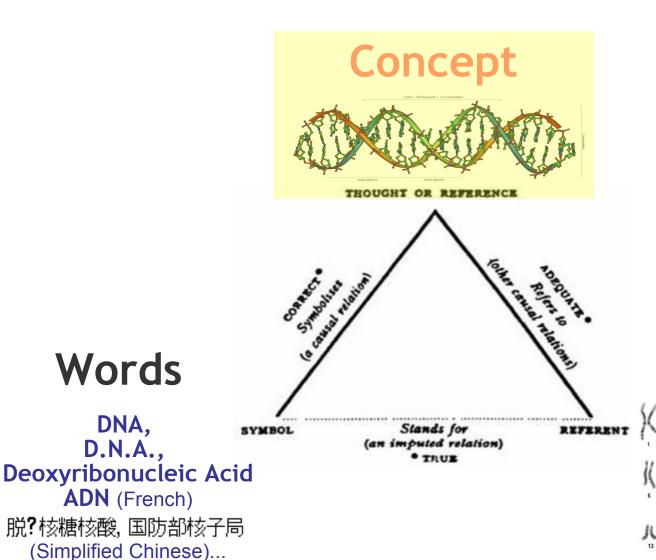University of Calgary

2008-02-11

Motivation & Goals

- Interoperability of maximum number of autonomous resources (databases & analysis programs)

- Minimal overhead of implementation
- Flexibility for bioinformaticians
- Usability for biologists

What are Semantics?

Concept
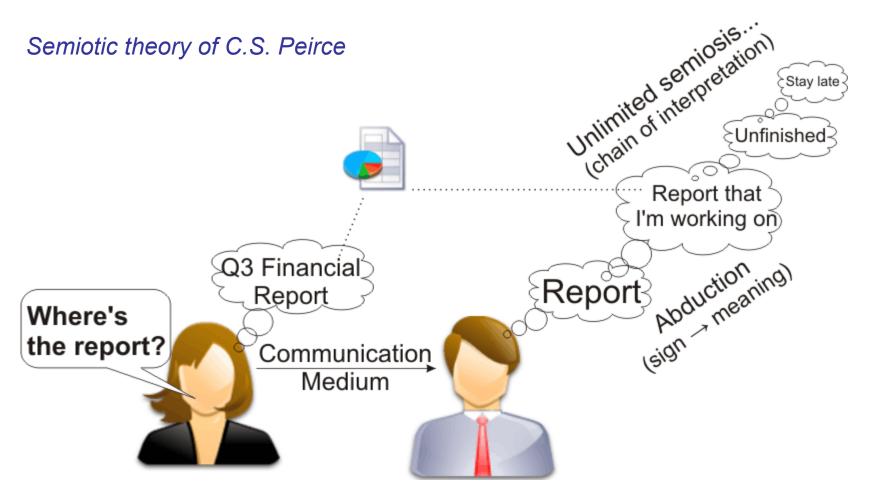
THOUGHT OR REFERENCE

CORRECT • Symbolises (a causal relation)

ADEQUATE • Refers to (other causal relations)

Words

Things

SYMBOL   *Stands for (an imputed relation)* • TRUE   REFERENT

**DNA,
D.N.A.,
Deoxyribonucleic Acid
ADN** (French)
脱?核糖核酸, 国防部核子局
(Simplified Chinese)...

Semantic Triangle from *The Meaning of Meaning* (1923)

by C. K. Ogden & I. A. Richards

Words have no fixed meaning in normal conversation

Semiotic theory of C.S. Peirce

Computers don't generally work this way, their interpretation of input is **strict**
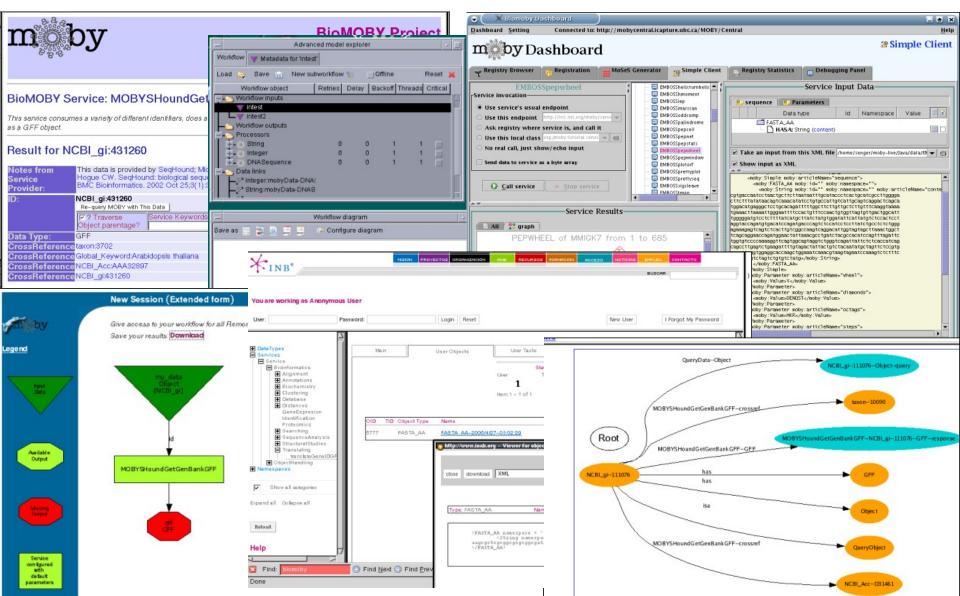
Paul's Pragmatic Maxims

- You cannot get rid of work

- Redistribute work amongst parties
    - Who's most motivated?
    - Who's most knowledgeable?

- Avoid repetition of work
    - Accessibility of code
    - Reusability of code

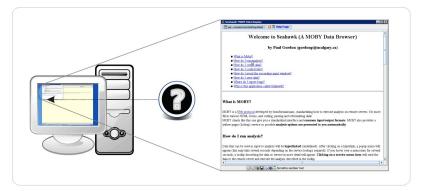# Many MOBY Clients…
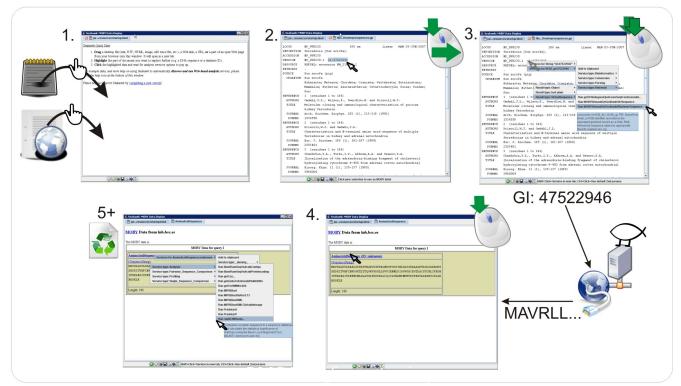
Seahawk: MOBY Client for Biologists

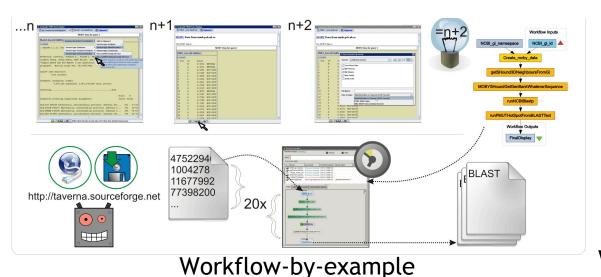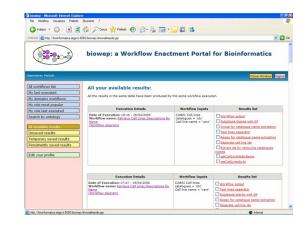## Drag 'n' drop, highlight existing data for use with MOBY Services

Service Provision Model

- Make existing computational resources available as MOBY services

- Build new platform tools as MOBY services

- Make it easier for biologists to create and run MOBY workflows for repetitive analysis

Workflow-by-example
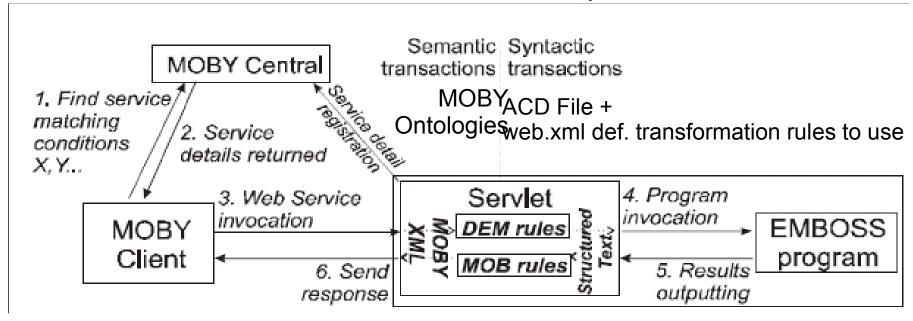
Workflow Execution Web Portals

MOBY Service Provision: Java

```java
import org.biomoby.shared.MobyDataType;
import org.biomoby.shared.data.*;
import org.biomoby.service.*;

@mobyService(name="ConvertAAtoFASTA_AA",
             type="FormatConversion",
             provider="moby.ucalgary.ca",
             author="gordonp@ucalgary.ca",
             in={"inseq:AminoAcidSequence"},
             out={"outseq:FASTA_AA"},
             description={"Converts amino acid objects into FastA formatted records,  ",
                          "primarily to increase inter-service compatibility "})

public class ConvertAAtoFASTA_AA extends MobyServlet{

    public void processRequest(MobyDataJob request, MobyDataJob result)
        throws Exception{

        // The input parameter for this method is registered as "inseq"
        MobyDataComposite aaSeqObj = (MobyDataComposite) request.get(" inseq");

        // SequenceString is a member of incoming AminoAcidSequence object
        MobyDataString aaStringObj = (MobyDataString) aaSeqObj.get(" SequenceString");
        String aaString = aaStringObject.toString();
        ...
```

MOBY Service Provision Legacy

Legacy non-XML apps (by MOBY Service provider)
EMBOSS = declarative "MOB" and "DEM" rules
+ web.xml + ACD file spec

Gordon P.M.K., Kawas E., Wilkinson M.D., Sensen C.W. (2007) Using a novel data transformation technique to provide the EMBOSS software suite as Semantic Web Services. To appear in Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (San Jose, California), November 2-4, 2007.

MOB Rules

```
CHARGE of HBB_HUMAN from 1 to 146: window 5

Position     Charge
1            0.100
2            -0.100
3            -0.400
4            -0.200
5            -0.200
6            -0.200
7            0.000
8            0.200
9            0.000
10           0.000
...
142          0.400
```

MOB rule (regex+XML)

```
<ProteinChargeTable articleName="" namespace="" id="">

  <String articleName="caption">CHARGE of HBB_HUMAN from 1 to 146: window 5</String>

  <tableRow articleName="header">
   <String articleName="cell">Position</String>
   <String articleName="cell">Charge</String>
  </tableRow>

  <tableRow articleName="row">
   <String articleName="cell">1</String>
   <String articleName="cell">0.100</String>
  </tableRow>
...
  <tableRow articleName="row">
   <String articleName="cell">142</String>
   <String articleName="cell">0.400</String>
  </tableRow>
</ProteinChargeTable>
```

MOBY XML

DEM Rules

MOBY XML

```
<DNASequence namespace="NCBI_gi" id="146411799">
  <String articleName="SequenceString" id="">TAGGCAT...GATGT</String>
  <String articleName="Length" id="">252</String>
</DNASequence>
```

DEM rule (XSLT)

```
>NCBI_gi|146411799
TAGGCATTACCCTACCAACTAACTAATGTTCCGCACCC
CCATTTTTAAGTGAAGCTGTGAAGCTCCTTTCTATTAC
...
GATGT
```
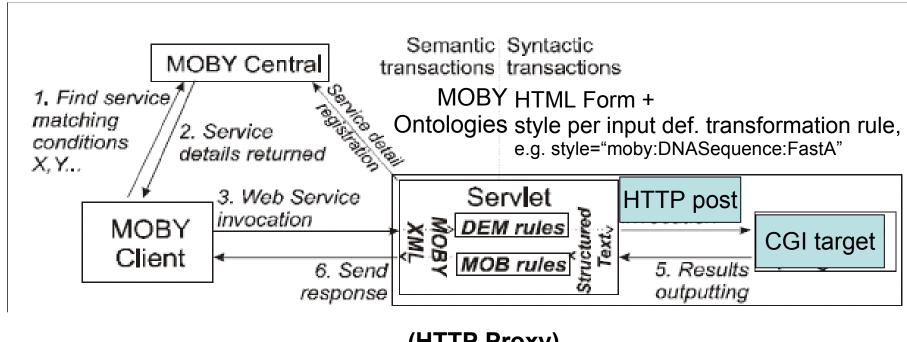
FASTA-formatted text

```
<xsl:template match="moby:GenericSequence"
    name="GenericSequence.1" mode="seqall">...
</xsl:template>
```

MOBY Service Provision
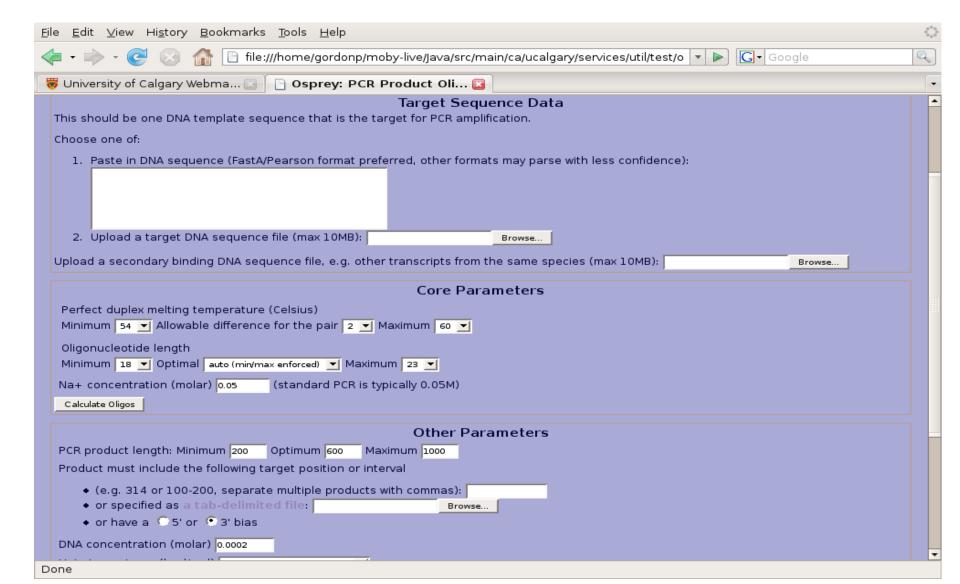
Legacy non-XML apps (by Web site authors)
HTML = declarative "MOB" and "DEM" rules
+ HTML form spec

Semantic transactions   Syntactic transactions

MOBY HTML Form +
Ontologies style per input def. transformation rule,
e.g. style="moby:DNASequence:FastA"

1. Find service matching conditions X, Y...

2. Service details returned

Service detail registration

3. Web Service invocation

6. Send response

MOBY Central

MOBY Client

Servlet
DEM rules
MOB rules

HTTP post

CGI target

5. Results outputting

(HTTP Proxy)

Paul Gordon PhD work

Wrapping CGIs

Semantic Markup via CSS

```html
<head>

    <meta name="moby:service"

        scheme="http://moby.ucalgary.ca/moby/MOBY-Central.pl"

        content="Primers,moby.ucalgary.ca,CalcPCRPrimers: Takes a DNA sequence

                and calculates PCR probes satisfying given conditions" />

    <meta name="moby:contact" content="gordonp@ucalgary.ca">

...
<form class="moby:CalcPCRPrimers:primers:Collection(DNASequence)" action="...

<textarea name="input" cols="61" rows="5" class="moby:inseq:DNASequence:fasta">

...<!-- existing selects, etc. -->

Na+ conc. (molar) <input type="text" size="7" name="salt_conc"

    value="0.05" class="moby::Float::[0,10]" alt="Na+ concentration (molar)"/>

PCR product: Minimum <input type="text" size="4" name="min_prod_len"

                value="200" class="moby:MinimumProductLength:Integer:100:[1,]"/>

...
```
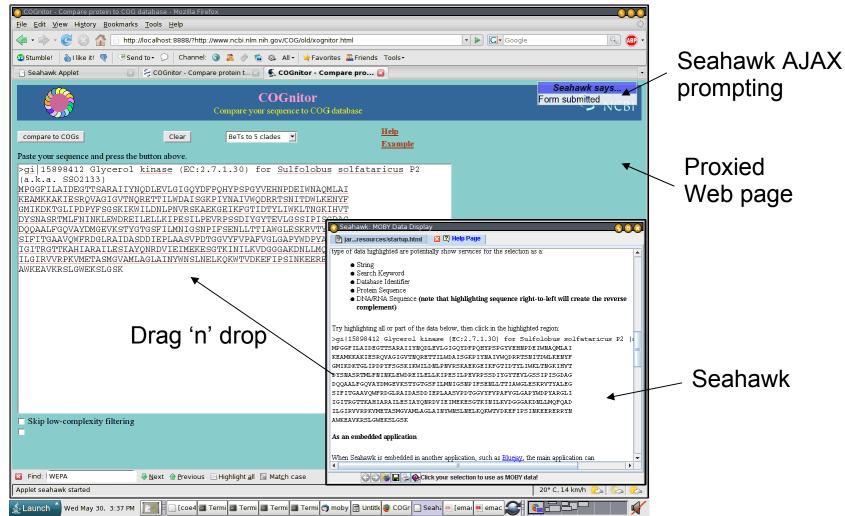
Next: Client-Server Convergence

Legacy non-XML apps (by end-users!)

HTML = declarative "MOB" and "DEM" rules
+ user drag 'n' drop in Seahawk

Seahawk AJAX prompting

Proxied Web page

Drag 'n' drop

Seahawk

Conclusions

- Semantics is enabling better "conversations" with the computer
- Distribute work amongst motivated participants
  - Ontology Building
  - Service Provision
  - Workflow Development
- Use a framework more general than you think you need, because you don't actually know what you need

Ω