

March 2006

Peter Ernst <P.Ernst@dkfz-heidelberg.de>

# Web services for bioinformatics applications



dkfz.

GERMAN  
CANCER RESEARCH CENTER  
IN THE HELMHOLTZ ASSOCIATION

- ***What are WebServices?***

- WebServices use SOAP (XML request/response messages) over HTTP
- WSDL-files (Web Service Description Language) describe the SOAP-service
- Clients read the WSDL
  - To select the proper connection settings, and
  - To choose a proper data-(de)serialization (computer language to XML mapping)

- EBI (European Bioinformatics Institute)



“The EBI has chosen to use the Web Services technology to expose its services in a programmatically accessible manner.”

*Nucleic Acids Research 2005 33(Web Server Issue):W25-W28, S. Pillai et al*

- NCBI (National Center for Biotechnology Information)



### Entrez Programming Utilities

Updated: February 17, 2006

- **EInfo**: Provides field index term counts, last update, ...
- **ESearch**: Searches and retrieves [primary IDs](#) ...
- **EPost**: Posts a file containing a list of [primary IDs](#) ...
- **ESummary**: Retrieves document summaries from a list ...
- **EFetch**: Retrieves records in the requested format ...
- **ELink**: links to Entrez databases or Related Articles
- **EGQuery**: Provides Entrez database counts in XML for a single [Global Query](#).
- **ESpell**: Retrieves spelling suggestions.
- [SOAP Interface for Entrez Utilities](#)
- [MEDLINE UI / PubMed ID Matcher](#)

<http://ws.apache.org/axis/java/>

http://ws.apache.org

Apache <Web Services > Project

Apache | Web Services | Axis

Search  Go

- **Axis**
  - Introduction
  - News
  - FAQ/Wiki
  - **Get Involved**
    - Overview
    - CVS Repository
    - Mailing Lists
    - Reference Library
    - Bugs
    - HowToBuildSite
  - **Axis (Java)**
    - Documentation
    - Releases
    - Installation
    - [User's Guide](#)
    - Developer's Guide
    - Integration Guide
    - [Architecture Guide](#)
    - Reference Guide
    - Reading Guide
    - Requirements
  - **Axis C++**
    - Latest Axis C++ Release!
    - Documentation
    - Download
    - Wiki Pages
    - Who we are
  - **Downloads**
    - Interim Drops

## Documentation

This is the documentation for [Apache Axis 1.2](#). If the version of Axis you are using is older or newer than this version, then this is the wrong documentation to be using. Read the version that came with your copy of Axis.

### Documentation for Axis Users

- [Installation Instructions](#)
- [User's Guide](#)
- [Client-side Axis](#)
- [Securing an Axis-based Web Service](#)
- [Axis Ant Tasks](#)
- [Reference Material](#)
- [Further Reading](#)

### Documentation for Axis Developers

- [API Documentation](#)
- [Building Axis](#) - Guidelines for building Axis with/without optional components.
- [Developer's Guide](#) - Collection of guidelines for developing code in Axis.
- [Integration Guide](#) - Description of APIs and development direction to allow integration into an existing web application server.
- [Architecture Guide](#) - Axis design concepts and rationale.

<http://www.cs.fsu.edu/~engelen/soap.html>

```
$ wsdl2h -o quote.h http://services.xmethods.net/soap/urn:xmethods-delayed-quotes.wsdl  
$ soapcpp2 quote.h
```

```
int main()  
{  
    net q; // "net" is the proxy class with a name that is the short name of the service  
    float r;  
    if (q.ns1__getQuote("IBM", r) == SOAP_OK)  
        std::cout << r << std::endl;  
    else  
        soap_print_fault(q.soap, stderr);  
    return 0;  
}
```

<http://soaplite.com/>

## 1.a. server

```
#!/perl -w

use SOAP::Transport::HTTP;

SOAP::Transport::HTTP::CGI
-> dispatch_to('Demo')
-> handle;

package Demo;

sub hi {
    return "hello, world";
}

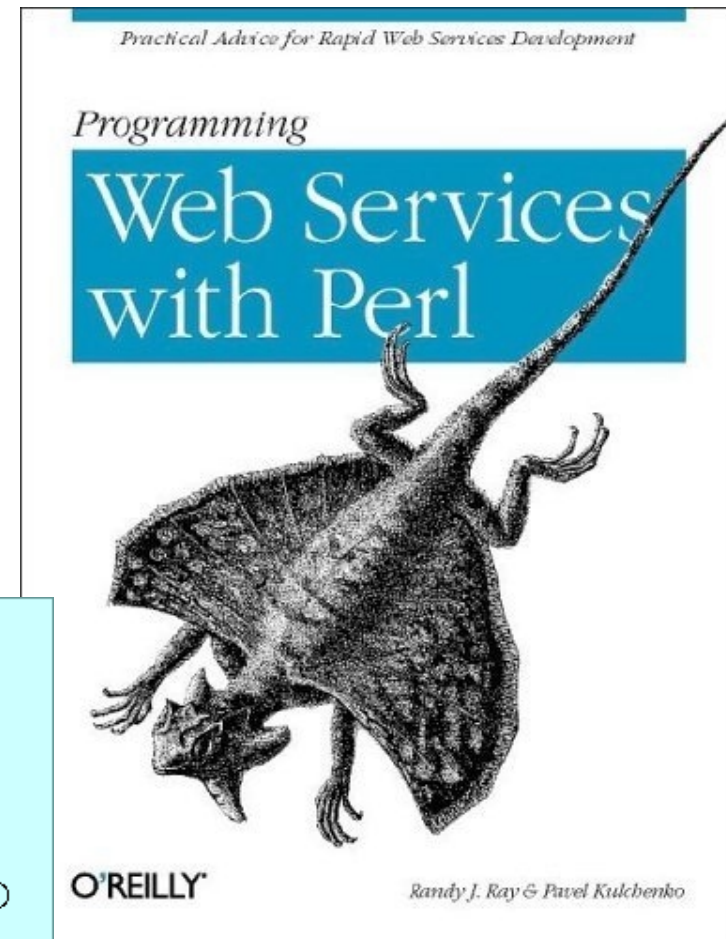
sub bye {
    return "goodbye, cruel world";
}
```

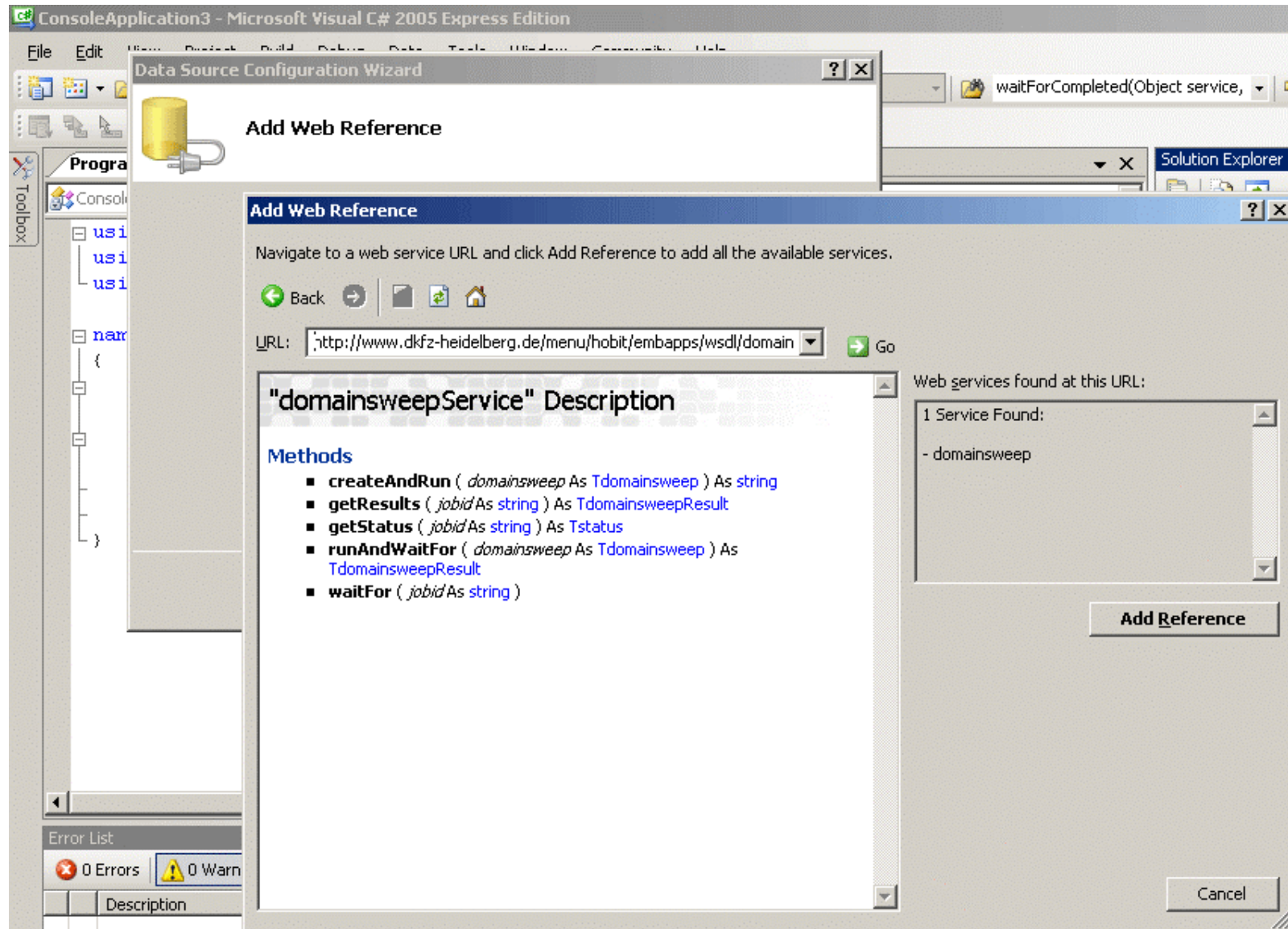
## 1.a. client

```
#!/perl -w

use SOAP::Lite;

print SOAP::Lite
-> uri('http://www.soaplite.com/Demo')
-> proxy('http://services.soaplite.com/hibye.cgi')
-> hi()
-> result;
```

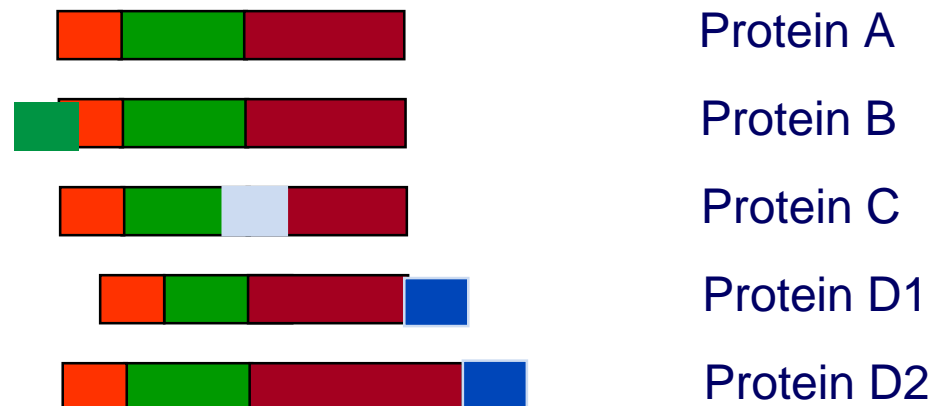




## Example: protein domains

- When dealing with a sequence of unknown function, the presence of similar domains can imply similar function.

Many genes and proteins are members of families which share a common biochemical function or evolutionary origin

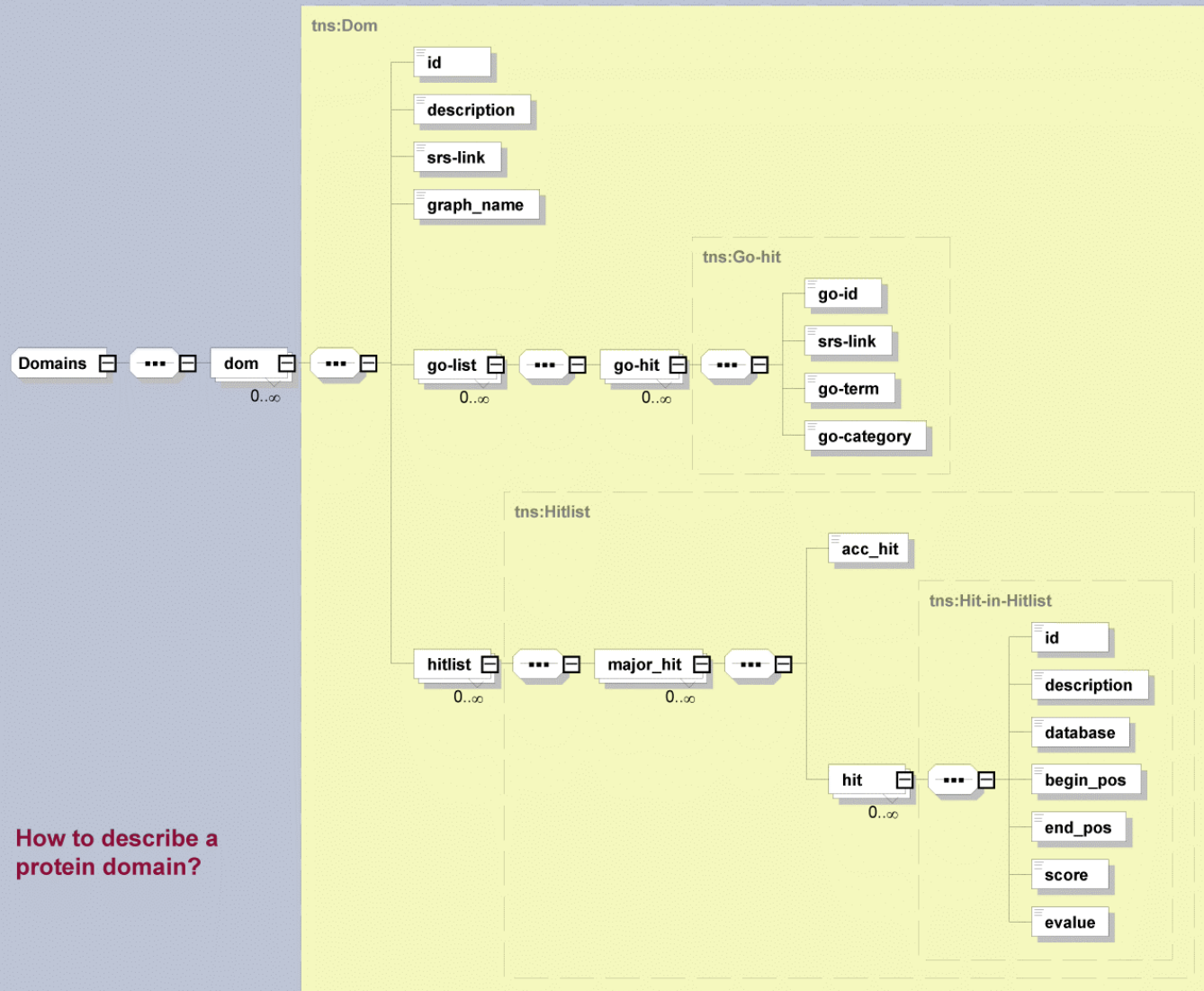


How to describe a protein domain?

There are many ways using XML to store the data.

Are there compatible ways to do this?

(e.g. like the “EBIApplicationResult” schema for Blast-like outputs)



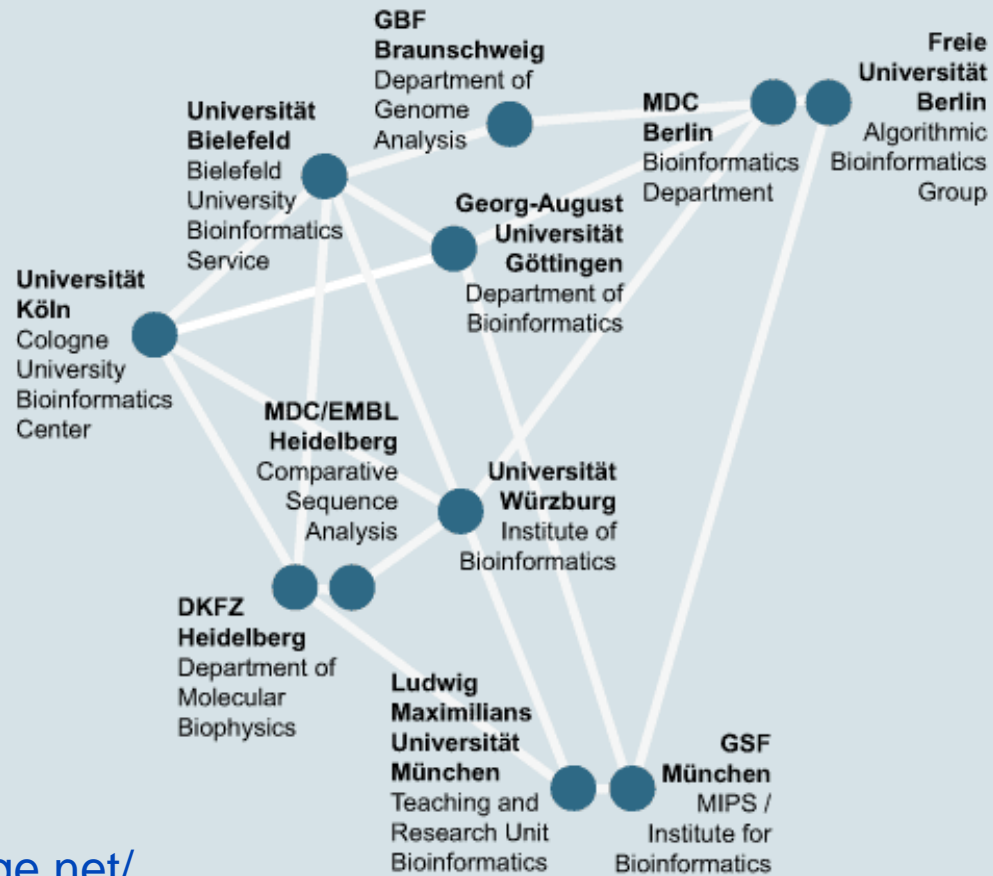
How to describe a protein domain?



- about hobit
- news
- downloads
- documentation
- development
- links
- contact

search

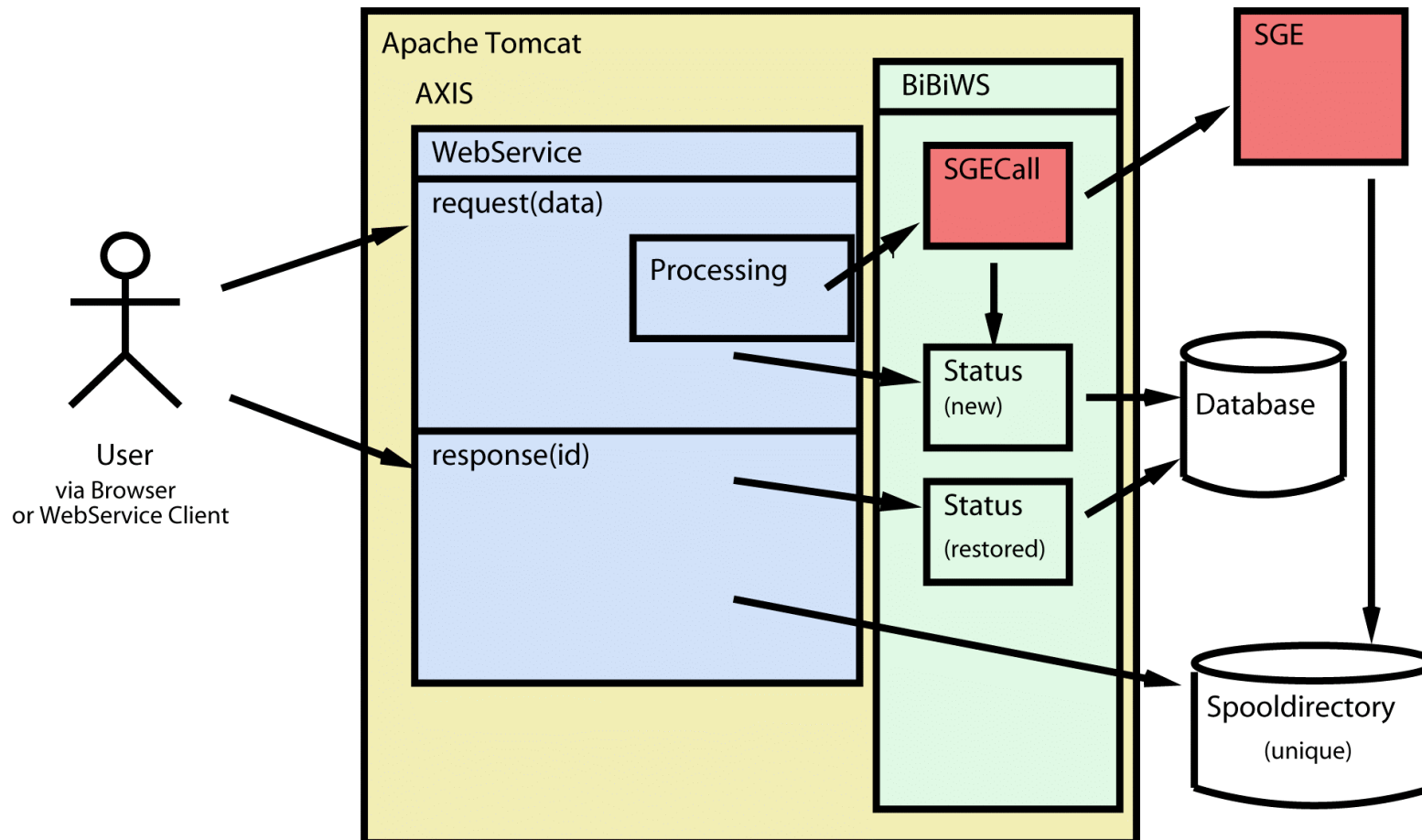
## helmholtz open bioinformatics technology



<http://hobit.sourceforge.net/>

# XML Schema: evolving standards for bioinformatics applications

Schema	Origin	Description
<a href="#">EBI FASTA BLAST</a>	<a href="#">EBI</a>	<p>The ApplicationResult.xsd used at the EBI for sequence similarity searches</p> <ul style="list-style-type: none"> <li>◆ Namespace: <a href="http://www.ebi.ac.uk/schema">http://www.ebi.ac.uk/schema</a></li> <li>◆ Comments: Used for BLAST FASTA</li> </ul>
<a href="#">GAME</a>	<a href="#">fruitfly.org</a>	<p>The Berkeley genome and sequence markup language.</p> <ul style="list-style-type: none"> <li>◆ Comments: Used in different Open Source Projects. Unfortunately only as DTD available.</li> </ul>
<a href="#">PSI-MI</a>	<a href="#">Proteomics Standards Initiative</a>	<p>The PSI MI format is a data exchange format for protein interactions.</p> <ul style="list-style-type: none"> <li>◆ Namespace: <code>net:sf:psidev:mi</code></li> <li>◆ Comments: Community standard. Supported by several important protein interaction databases.</li> </ul>
<a href="#">MAGE-ML</a>	<a href="#">MAGE</a>	<p>The standard for the representation of microarray expression data facilitating the exchange of microarray information between different data systems.</p> <ul style="list-style-type: none"> <li>◆ Comments: Community standard. Unfortunately a DTD.</li> </ul>
<a href="#">SBML</a>	<a href="#">SBML</a>	<p>The Systems Biology Markup Language (SBML) is a computer-readable format for representing models of biochemical reaction networks.</p> <ul style="list-style-type: none"> <li>◆ Namespace: <a href="http://www.sbml.org/sbml/level2">http://www.sbml.org/sbml/level2</a></li> <li>◆ Comments: Supported by over 80 software systems in systems biology.</li> </ul>
<a href="#">UniProt XSD</a>	<a href="#">UniProt / EBI</a>	<p>The UniProt for protein database format.</p> <ul style="list-style-type: none"> <li>◆ Namespace: <a href="http://uniprot.org/uniprot">http://uniprot.org/uniprot</a></li> <li>◆ Comments: Successor of the SwissProt SP-ML format.</li> </ul>



Webservice setup Bioinformatics University Bielefeld, Germany (BiBiServ)  
(from diploma thesis: Henning Mersch, 2004)

- WebServices are
  - verbose (XML)
  - firewall-friendly (HTTP)
  - supported
    - in many programming languages
    - by big industrial players (Microsoft, IBM, etc.)
    - by important bioinformatics institutes
  - easy to use in clients
- XML Schemas:
  - provide the tool to standardize data exchange between bioinformatics peers
  - because of its machine-readability, automatic data translation between exchange- and locally used format is possible (toolkits)