BioCatalogue
The Life Science Web Service Registry

Eric Nzuobontane
European Bioinformatics Institute

EBI is an Outstation of the European Molecular Biology Laboratory.
Web Services(1)

- Programmatic Interfaces to Services.
- Standard interfaces.
  - HTTP, XML
- Two big families:
  - SOAP and REST.
- Three main roles
  - Provider, Consumer, Registry
Web Services(2) (as building blocks)

Software Lego™ that works across the web and underpins enterprise SOA.
Web Services(3)

- Software system designed to support **interoperable machine-to-machine interaction over a network**

![Diagram of machine-machine communication](image-url)
Web Service Orchestration

- Accomplish complex business processes
  - Use multiple services
  - Reduced cost
  - Standardized communication interfaces & protocols

- Workflow (service orchestration) engines
  - Visualize business process composition

- Some engines used in life science
  - Pipeline Pilot™, VIBE®
  - Taverna, Kepler, Triana, MIGenAS,
Transmembrane and signal peptide prediction using three methods (services):

- **EMBOSS tmap with a single sequence**: Uses Soaplab tmap.
- **Phobius**: Uses EBI's WSPhobius web service.
- **TMHMM and SignalP**: Uses the TMHMM and SignalP methods of InterProScan via the EBI's WSInterProScan service.

The results of the three methods are converted into GFF format and collated.

http://www.myexperiment.org/workflows/216.html
Web Services in Life Science

- Providers
  - European Bioinformatics Institute (EBI)
    - http://www.ebi.ac.uk/Tools/webservices/
  - National Centre for Biotechnology Information (NCBI)
  - DNA Databank of Japan (DDBJ)
    - http://www.ddbj.nig.ac.jp/
  - Commercial & other providers
Searching for web services

- Search engines
  - Google, bing, yahoo…

- Web service search engines
  - Generic
    - seekda, xmethods, service finder…
  - Domain Specific
    - BioCatalogue, DAS, Biomoby

- Providers websites
  - EMBL-EBI, DDBJ, NCBI

- Ask someone!
Publishing web services

- **Company website**
  - Does that have max exposure?
- **Search engines**
  - Are keyword search good enough?
  - Search engines designed for web page, not applications
- **Journals**
  - Selected audience
Understanding/Using web services

- What the service really does(?)
  - Interface show syntax (wsdl)
  - Function + operational info needed

- Sample clients
  - Any available for immediate use

- Constraints
  - Conditions of use of service
Service Properties

- **Different classes of consumers**
  - Systemic, non-systemic

- **Varying importance (for different users)**
  - Availability
  - Reliability
  - Consistency

- **Updates**
  - Latest updates
  - Specific revisions
Four Major Problems

• Web Services are hard to find
  • Where and how do I find them?

• My web services are not visible
  • Where do I publish my web services to achieve maximum visibility?

• Web services can be volatile (life cycle)
  • What is the status of the services I am interested in?

• Web Services can be poorly describe
  • Is there sufficient information available to allow proper use of service?
BioCatalogue
The Life Science Web Service Registry

http://www.biocatalogue.org

An Open, Public, Curated Catalogue of Web Services in Life Sciences

Launched – June 2009

Publication
BioCatalogue’s Mission

"Web Services are hard to find"

DISCOVER
- Find the right Web Service
- Powerful search and filtering
- Information from providers and community

"My Web Services are not visible"

REGISTER
- Easily register Web Services
- Instantly available to everyone
- Providers can advertise, describe and monitor their Services

"Web Services are poorly described"

ANNOTATE
- Anyone can describe and annotate
- Ongoing expert curation
- Social curation by the community

"Web Services are volatile"

MONITOR
- Services change and get outdated
- BioCatalogue monitors Services
- Monitors availability and reliability
Service Search(1)

- **Keyword search**
  - Google style simple keyword search

- **Browse by tags**
  - Community and ontology tags

- **Navigate by service categories**
  - Categorization of services by providers/curators/users

- **Navigate by service providers**
  - Service grouped by their institutes/organization
Service Search(2)
Service Registration (Publish)

- Simple registration process
  - Only a URL is required

- Register different types of Services
  - SOAP, REST

- Register batch of services
  - Soaplab

- Instant availability
Service Annotation

- Provider & Community annotation
  - Providers as ‘authority’ on services
  - Community in addition to providers
- Improve search results
  - Users can easily find service
- Improved understandability
  - Documentation for your service
- Information on constraints
  - Fair use policy
Position Specific Iterative BLAST (PSI-BLAST) refers to a feature of NCBI BLAST 2.0 in which a position-specific scoring matrix (PSSM) profile is constructed from set of BLAST alignments. The PSSM is then used to search the database in subsequent iterations allowing the detection of distant evolutionary relationships.
Programmatic access(1)

- REST API available
  http://www.biocatalogue.org/wiki/public

- Integrate in your tool
  - Taverna, NCRI
  - http://www.ncri-onix.org.uk/portal/#S103a

- Query programmatically
  - Eg service status
Setup Private instance(1)
Setup Private instance(2)

- Source code available for free
  - http://rubyforge.org/projects/biocatalogue/

- Documentation on installation

- Help from developers on setting up
  - Join the BioCatalogue friends
  - Updates announced and made available for free
Questions?

Sign up to BioCatalogue

http://www.biocatalogue.org

Contact us

contact@biocatalogue.org
References

- http://www.biocatalogue.org
- http://www.ebi.ac.uk
- http://www.myexperiment.org
- http://www.w3.org/TR/soap/
- http://energybenchmarking.lbl.gov/aob.html