

# A BioCatalogue

## Cataloguing Web Services for the Life Science Community

**Carole Goble**, Khalid Belhajjame, Robert Stevens, Jiten Bhagat, Franck Tanoh,



open middleware  
infrastructure  
institute uk

Katy Wolstencroft, Steve Pettifer

University of Manchester, UK



EMBL-EBI



Rodrigo Lopez, Thomas Laurent,

Hamish McWilliams, Eric Nzuobontane

European Bioinformatics Institute, UK

UNIVERSITY OF  
**Southampton**

School of Electronics  
and Computer Science

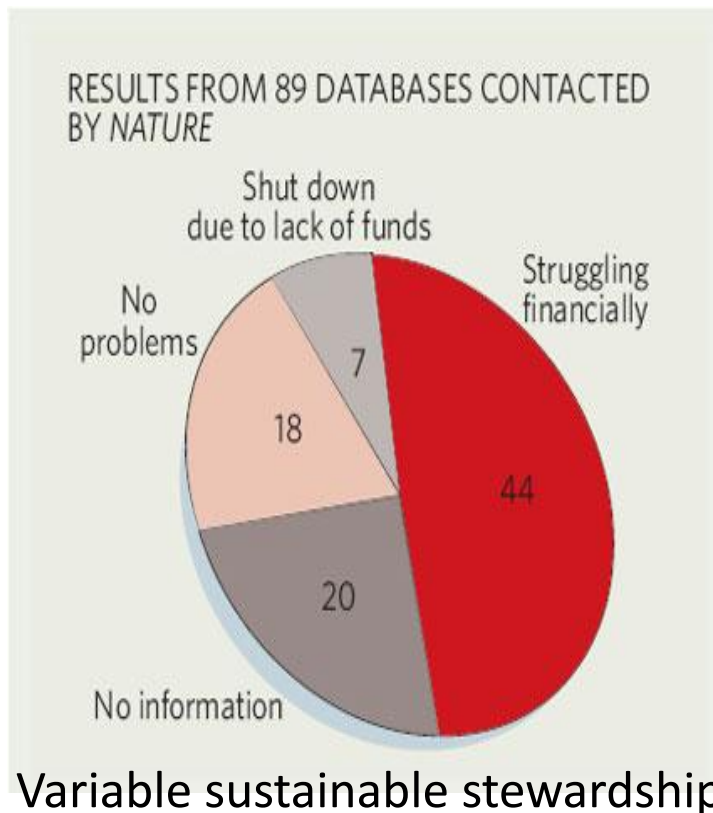
*David De Roure, myExperiment*

# Web Services in the Life Sciences

- Programmatic Interfaces to services on the rise
- EMBL-European Bioinformatics Institute
  - 3 million/month accesses to Web Service APIs
  - 1 million/month compute jobs > 50% are over WS
- Guessimate 1000-1500 services.
- Why?
  - Specialisation and segregation of methods from monolithic servers.
  - How one should publish data.
  - Automated Life Science applications, like workflow systems - Taverna, Kepler, Triana, Trident, KNIME, BPEL .....

# Chain stores and Boutiques

- Major data centres and national centres
  - EMBL-EBI (UK), DDBJ, PDBJ (Japan), NCBI, SDSC PDB (USA)
- Investigator and community projects



- Kanehisa Laboratory, Kyoto, Japan
- BASIS, University of Newcastle, UK
- Biomolecular Interaction Network Database, BIND, University of Toronto, Canada
- Institute of Bioinformatics, Tsinghua University, China
- EMAP, Edinburgh Mouse Atlas Project, UK
- The Chemical Informatics and Cyberinfrastructure Collaboratory (CICC), Indiana University, USA

and more and more....

# Service Flavours

- Generalist
  - SOAP
  - REST
- Specialist
  - DAS (Distributed Annotation Services)  
*[www.biodas.org](http://www.biodas.org)*
  - BioMOBY  
*[www.biomoby.org](http://www.biomoby.org)*



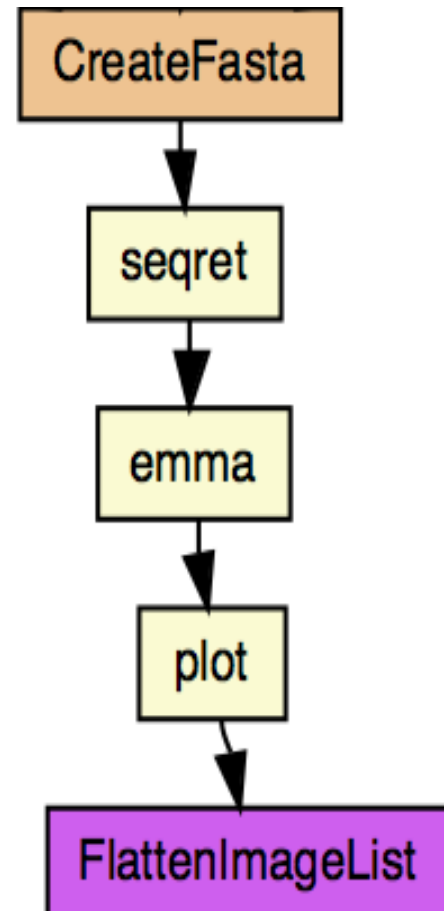
# Web Services in the Wild

## Visible? Findable?

- “EMMA” is the Clustalw multiple sequence alignment program from the Emboss suite
- Poor adoption for providers.
- Forum for advertising and shopping.

## Executable?

- WSDL, WADL, WSDL2, Other kinds of services.
- Transcend the specific grounding



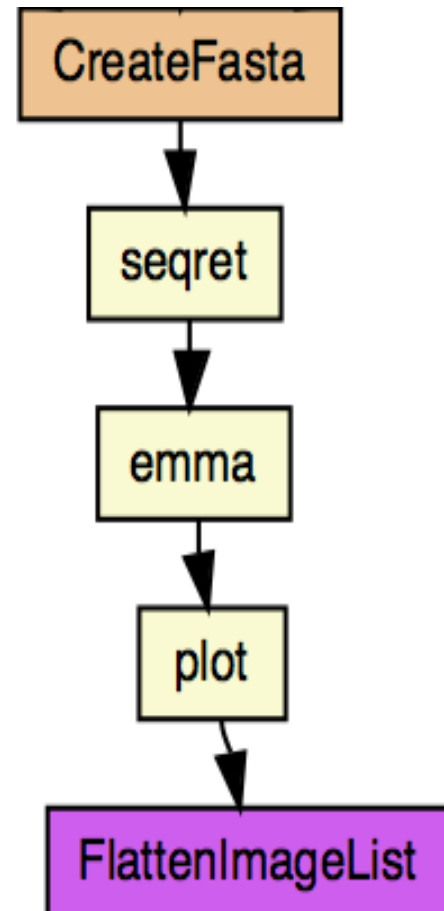
# Web Services in the Wild

## Understandable?

- Input0:string, Output0: string?
- What does the SeqRet actually do?
- Examples? Example data? Example Parameter configurations? Input-Output correlations?
- Adequate documentation for anonymous reuse.

## Usable? Available?

- Quality of Service, robustness, test scripts?
- Stability and dependability (see BioMART)?
- Licensing, execution restrictions?
- Trust and risk.
- Monitoring and intelligence gathering.



# Cataloguing Services

- Investigator and project specific registries
  - EMBRACE, BioSapien, Stargate Portal
- Community lists
  - Bioinformatics Links Directory, BioLinks, BioPlanet,
- Project specialist registries
  - BioMOBY Central, DAS Registry, myGrid Registry, Sswap
- General catalogues and search engines
  - SeekDa!, Web Services List, XMethods

**Sustainability  
and curation**

**Accessibility**

**Rich  
annotation &  
customisation**

**Provider  
engagement**

# Lets Pool our Knowledge

- A reliable, trusted, up to date and sustained catalogue customised for the Life Sciences.
  - EBI curation and service commitment
- Discovery interface for decision support.
  - Drawing on myExperiment and EBI legacies
- Community and specialist curation.
  - Pooled and accumulative annotation.
  - A platform for service monitoring and analytics.
- Incorporated into applications and mashups.
  - Itself a web service, with a (REST) API.



## BioCatalogue: providing a curated catalogue of Life Science Web Services.



BioCatalogue will provide a **single registration point** for Web Service providers and a **single search site** for scientists and developers.

BioCatalogue will also act as a place where the **community** can **find contacts** and **meet** the experts and maintainers of these services.

**Started June 08**

**Closed pilot Dec 08**

**Pilot release April 09**

**BioCatalogue-Friends focus group**

**Perpetual beta**

**Three year award**



"Web Services are hard to find..."

### SEARCH

Scientist, tool developers, bioinformaticians will be able to find the right Web Service they were looking for, thanks to an **easy and powerful search interface** harvesting the information made available by the Web Services providers and the BioCatalogue community.

"My Web Services are not visible..."

### REGISTER

**Service providers** will be able to **easily register** their Web Services in the BioCatalogue, making them **instantly available** to the scientific community as well as the tool developers.

"Web Services are poorly described..."

### ANNOTATE

**Expert curators** will provide oversight, monitor the catalogue and provide **high quality annotations** for services. The wider **community** will also participate to this effort using **social networking** for recommending, tagging, commenting and rating the services.

"Web Services are volatile..."

### MONITOR

Web Services are **volatile**. They change their location, capability and interaction or become **outdated**. BioCatalogue will allow agents to **monitor** the Web Services and **automatically** add information to the catalogue.

# Influences

Home » Workflows » Search Results

Didn't find what you need?



**Broadband Running Slow?**  
It Might Be Your Computer. Free Scan.  
Gold Award From PC Advisor.

**Speed-up your PC now**  
Speed-up your Computer in 2 min. Free  
Download - 100% Guaranteed.

Ads by Google

Login | Register

3 workflows found for "cag"

Note: some items may not be v

Taverna 1

Original Uploader

Credits: 0

License: C

Wei Tan

Rating: 0.0

## News & Info

- Home / News
- How Broadband Works
- Our Guides
- FAQ
- Videos
- Router Reviews

## Service Providers

- Broadband Provider List
- ISP Search
- Compare Broadband
- Rate Your Provider
- Resolving a Problem
- Availability Check

## Discussion

- Forums

## thinkbroadband

- Speed Test
- Tools
- Links
- About

Ads by Google

## Service Provider Comparison

Your selected ISPs have been compared below. Some ISPs may not appear every month, this is because there is not enough information to produce a fair and accurate figure. More detailed information is displayed if you hover your mouse over the bars in the chart below.

U.S. Department of Health & Human Services

HHS.gov

Improving the health, safety, and well-being of America

Frequent Questions

Search

**Hospital Compare** - A quality tool for adults, including people with Medicare

Hospital Compare Home Help Glossary Resources

<< Back to Choose Hospitals to Compare

Medicare.gov | Use Larger Font | E-mail This Page

## Compare Hospitals

Below are the hospital(s) y

**Your Selected Hospital**

**Hospital Process**

**Surgical Care Imp**

Hospitals can reduce right time on the day about how to preven

VivaLaDiva.com  
A WOMAN'S RIGHT TO SHOES

SIGN IN  
MY ACCOUNT  
REGISTER

Incentives  
Help  
Contact Us

Search  
Keyword or Item Number

All Categories



YOUR BAG: 0 ITEMS

Shopping Bag  
Go To Checkout  
Wish List  
Quick Order

BRANDS | STYLES | NEW IN | HANDBAGS | ACCESSORIES | SALE | JEWELLERY

Refine your results by

High Heels

- Ankle Strap
- Court
- Formal Court
- Mary Jane
- Mule
- Peep Toe
- Platform
- Sandal
- Sling Back
- Stiletto
- T-Bar
- Wedge

Shoe Size

2	6
2 1/2	6 1/2
3	7
3 1/2	7 1/2
4	8
4 1/2	9
5	10

Home >> Styles >> High Heels

**High Heels** - Stunning designer shoes with towering high heels

Give yourself the ultimate lift in a pair of sky-high heels, from formal court shoes to sexy stilettos, pretty peep toe slingbacks to wicked wedge sandals.

824 Items found [Show 15 30 45 per page]

Sort by: Our choice

1 2 3 4 5 ... 55 Next >



Strutt Couture Shino Platform  
Save £115.00  
Now £115.00



Schuh Venchy Brooch Sandal  
£53.62



Irregular Choice Cortesana Court  
£63.37

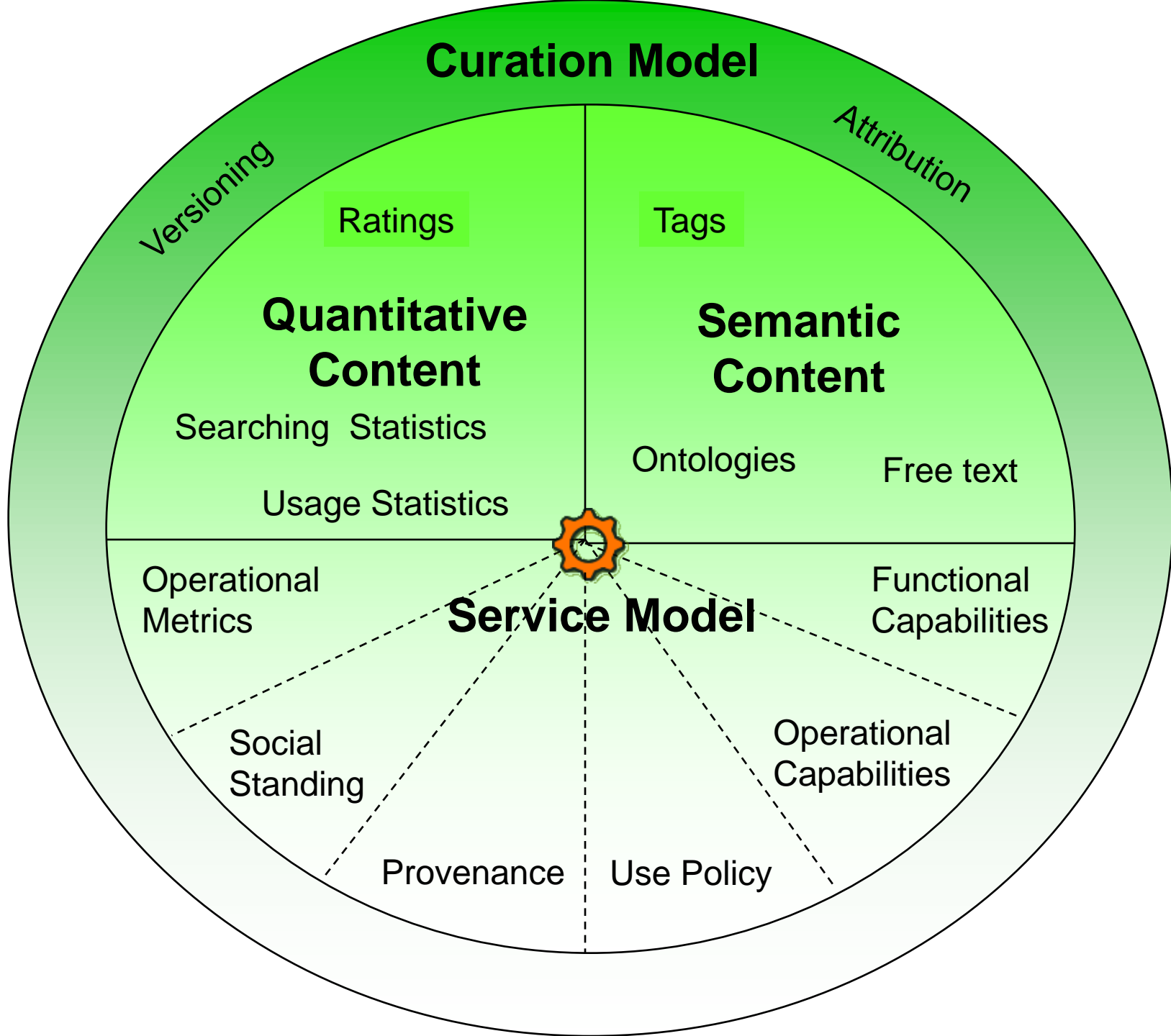


Harlot Court Shoe  
£77.99



Harlot Rome Mary Jane  
£77.99

# Service Profile Wheel



# Discovery

Search  
Browse/Shop



# Sorting Ranking

Matchmaking  
Customised



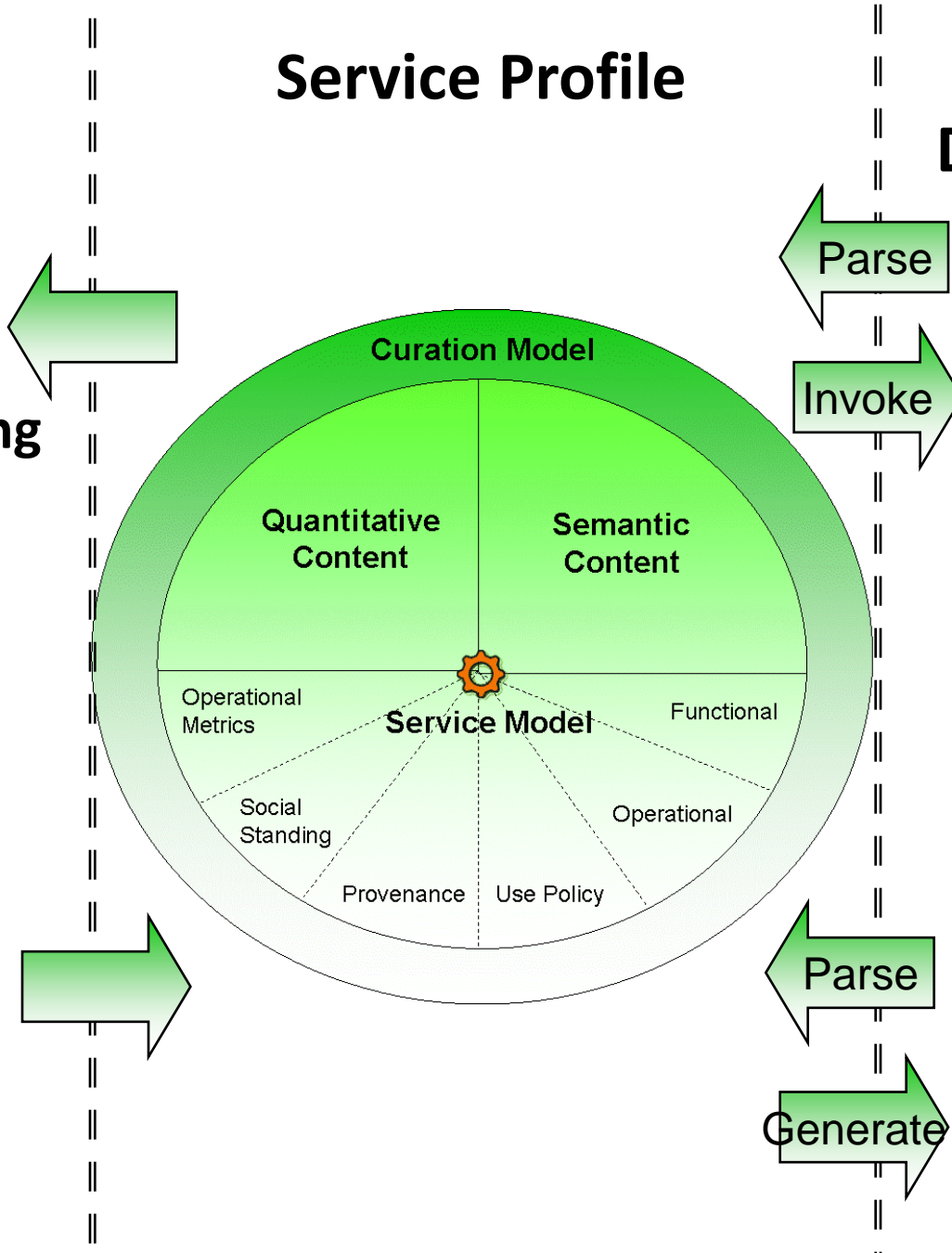
# Analytics

Searches  
Services Profiles  
Workflows



# Monitoring Validating

## Service Profile



## External Descriptions

Parse

Invoke

WSDL

WADL

WSDL2

SAWSDL

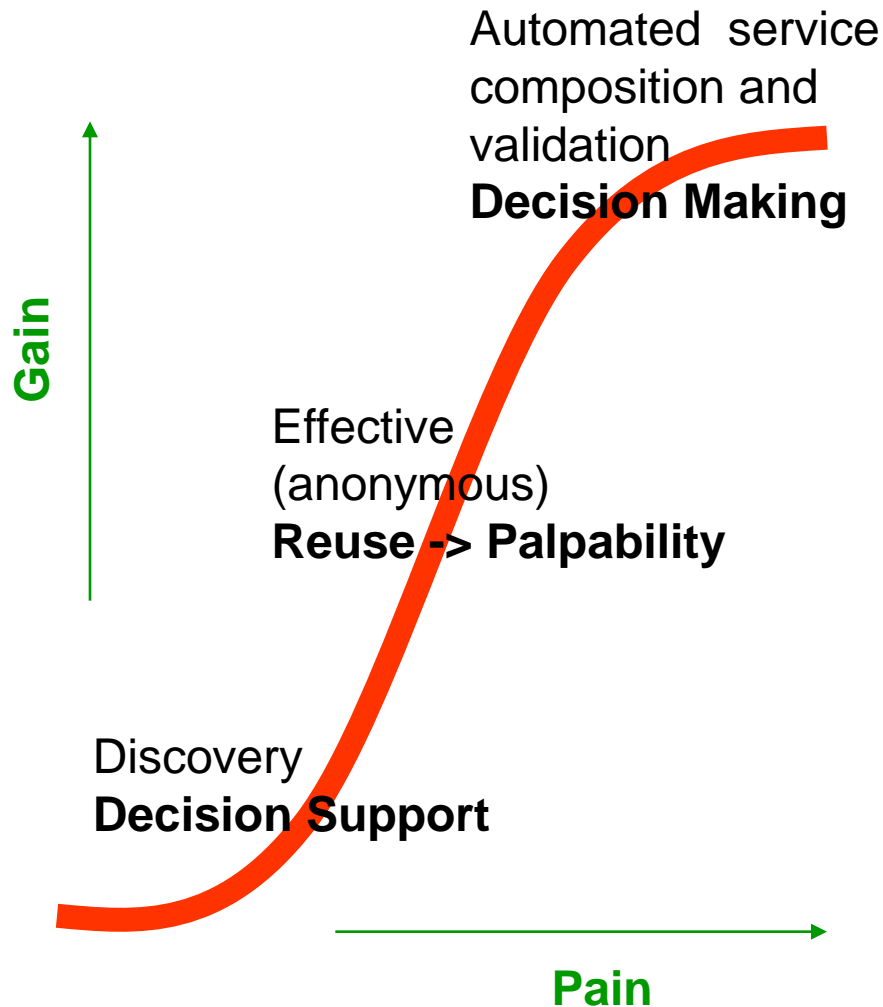
A.N.  
Other

SA-REST

Parse

Generate

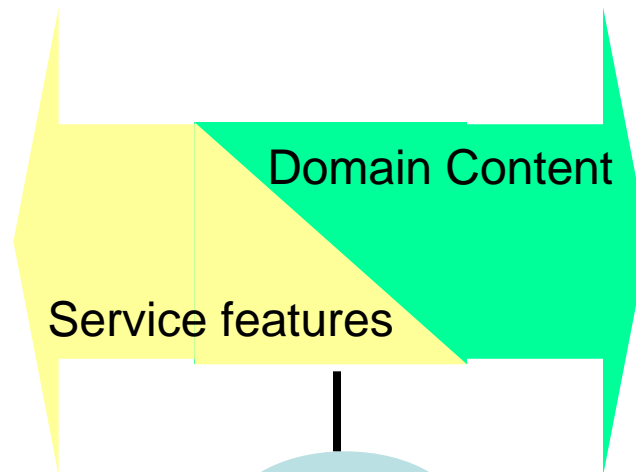
# Modelling Functional Capability



- WSMO  
<http://www.wsmo.org>
- OWL-S  
<http://www.w3.org/Submission/OWL-S>
- SAWSDL  
<http://www.w3.org/2002/ws/sawSDL/>
- .....
- Tags
- Ontology
- myGrid Service Ontology
- Text Descriptions

# myGrid Functional Capability Ontology

**Service**  
**Operations**  
**Inputs**  
**Outputs**  
**Task**  
**Method**  
**Resource**



**Informatics**  
**Bioinformatics**  
**Molecular Biology**  
**Formats**  
**Tasks**

Grounding

WSDL

W3C OWL and RDFS  
Number of classes ~750  
myGrid and BioMOBY

Free text and tagging  
in the user's language

Smart interfaces  
for people

Semantically annotated services for  
driving interfaces and automated  
processing



# People-Powered Registration

- By Provider and by Proxy.
- Ownership.
- Incentives
- Completeness vs Cost.
- Relative rankings feedback.
- Visibility and reputation.  
(which may not always be  
flattering)
- Do not presume that providers  
are unhelpful.





# People-Powered Curation

- Third party and Provider
- Curation@Source/Delivery
- Incentives.
  - Quick and easy.
  - Credit (and Blame).
- Incremental and partial descriptions.
- Peer review. The Wisdom of the Wisdom of the Crowd
  - Quality, Slander
- Content.



Distributed Human Grid of Annotators. Annotation Jamborees. T Shirts.

# Expert Curation

- Added value of Biocatalogue
  - Review
  - Quality assurance and Trust
- Enriched annotations
- A curation pipeline.
  - Tags to Ontologies.
  - Ontology husbandry
- A Sweatshop.
  - How do we make this smarter?

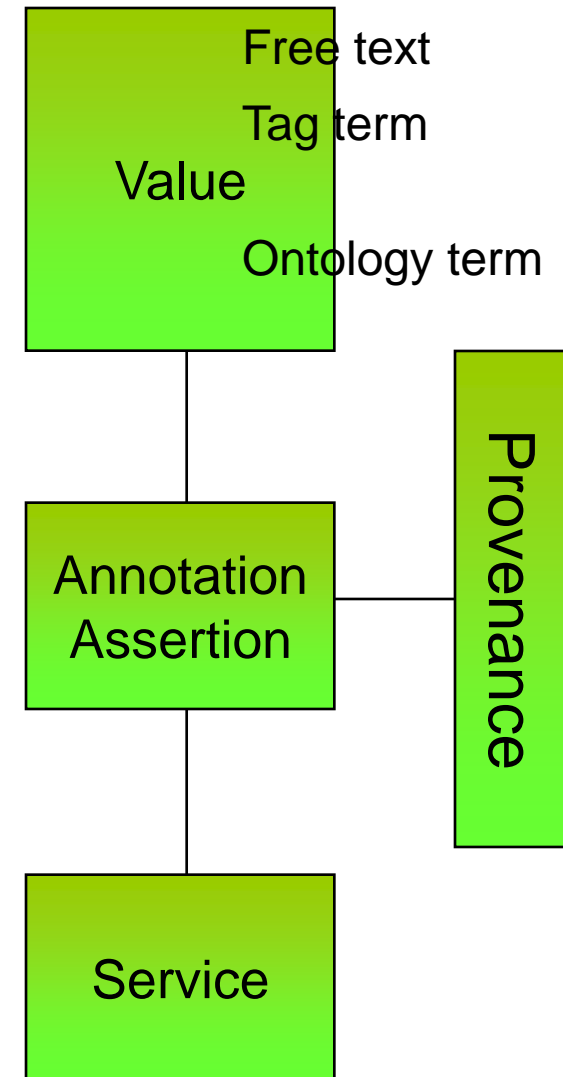


# Uniform Annotation model

- Minimum for discovery and invocation
- Partial annotations
- Multiple annotations
- Polymorphic: text, tags, statistics, ontologies

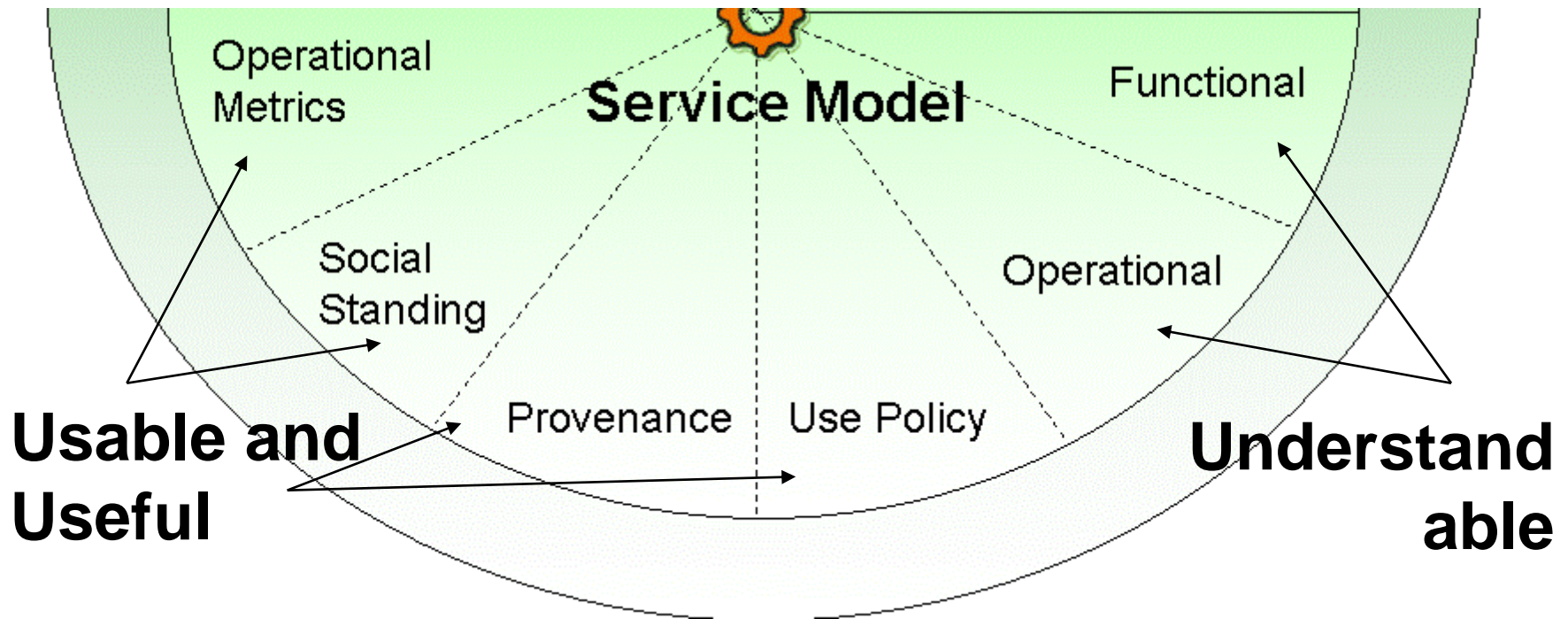
- Annotation provenance
- Trust
- Curation pipeline and monitoring

- Multiple providers
- Multiple versions
- Multiple deployments



# Ranking, Sorting, Filtering and Comparing

- Grading: bronze -> platinum
- Presence, quantity and quality
- Judgement by the users, not us.



## Search Results

results 1 to 10 of 72

sort by

- relevance (default)
- relevance (default)
- available documentation
- availability last week
- availability last year
- provider
- country
- age



**StretcherService**

use now



by [vt.edu](#)

Stretcher finds the best global **alignment** between two **sequences**. see:

[http://staff.vbi.vt.edu/pathport/services/#align\\_10](http://staff.vbi.vt.edu/pathport/services/#align_10) (community description)

Tags: [similarity](#), [alignment](#), [search](#), [sequence](#), [global](#), [calculation](#), [university](#), [free](#), [linear space](#)



**ContigService**

use now



by [vt.edu](#)

Creates assembled contigs from trace files. see:

### Related Searches

Other users who searched for "sequence alignment" searched also for:

[logo](#) [protein](#) [gene](#) [ontology](#)  
[seqlogo](#)

Advanced Search

Services Tag Cloud

Providers by Countries

Most Used Services

My Bookmarks

Recently Found Services

Latest service updates

**RNAmmer**

5 hours ago  
status changed to  
WARNING



**oligo\_analysis**

8 hours ago  
status changed to  
FAILED



**purge\_seq**

8 hours ago  
status changed to  
FAILED



**matrix\_scan**

8 hours ago  
status changed to  
FAILED



**retrieve\_seq**

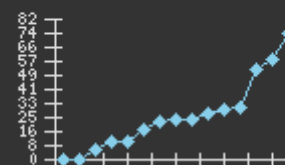
8 hours ago  
status changed to  
FAILED



### Current Statistics

Total Users: 48

Total Services: 75



Public News Announcements New Items Updated Items User Opinions

### Latest Tags

**d-grid** on [Workflow Pattern - Structured Loop \(C\)](#) by **Andreas Hoheisel** (4 days ago)

**example** on [Workflow Pattern - Structured Loop \(C\)](#) by **Andreas Hoheisel** (4 days ago)

**gwes** on [Workflow Pattern - Structured Loop \(C\)](#) by **Andreas Hoheisel** (4 days ago)

**gworkflowdl** on [Workflow Pattern - Structured Loop \(C\)](#) by **Andreas Hoheisel** (4 days ago)

### New Items

[File: eLab architecture breakout group notes](#) by **Sergey** (2 days ago)

[Workflow: Workflow Pattern - Structured Loop \(C\)](#) by **Andreas Hoheisel** (4 days ago)

[Workflow: Workflow Pattern - Structured Loop \(B\)](#) by **Andreas Hoheisel** (4 days ago)

[Workflow: Workflow Pattern - Structured Loop](#) by **Andreas Hoheisel** (7 days ago)

[Workflow: blastp using the MRS system](#) by **Bas Vroling** (7 days ago)

[File: Popfly Blocks for Protein Characteristics and Prediction Mashup 3](#) by **Felicity** (days ago)

[File: Popfly Blocks for Protein Characteristics and Prediction Mashup 2](#) by **Felicity**

### Bookmark & Share

© Add This

Favorites

Del.icio.us

Digg

Google

MySpace

Facebook

Reddit

Live

Furl

Yahoo MyWeb

StumbleUpon

More...

# Auto Curation



## Auto scavenging

- SeekDa!

## Auto Annotation

- Specialist parsing
- Auto-tagging
- Text mining
- Inferring service descriptions from myExperiment workflows (Quasar framework)

## Auto Monitoring

- Test Workflows / scripts
- Service monitoring
- Feeds from applications and third parties: dial home diagnostics, customer reports, predicted down times

## Auto Usage Analytics

- Workflow usage
- Search patterns

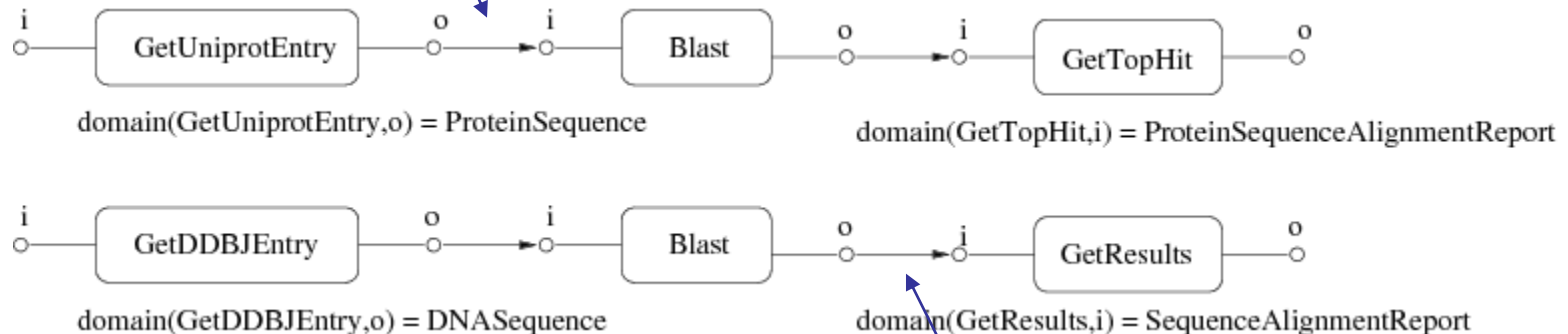


# Quasar

Quality Assurance of Semantic Annotations for Services

Using mismatch-free workflows to infer information about the semantics of linked parameters

$$(ProteinSequence \cup DNASequence) \subseteq domain(Blast, i)$$



$$domain(Blast, o) \subseteq ProteinSequenceAlignmentReport$$

Users	registration ownership	Identity management account management	profile management bookmarking	notification
Services	soap services scavenging	registration dashboard wsdl parsing	versions instances	Identity management
Discovery	text search tag search	browse and drill down sorting on criteria and categories	usage-based recommendations	
Curation	tagging seeded	specialist parsers. controlled vocab.	ratings recommendations.	500 services 250 full curated
Monitoring	registration test scripts WsdI monitoring	live tests	QoS app feeds	
Integration	REST API Open Search	myExperiment		
Content	Batch migration	Policy identification Provider engagement		

**Pilot**



Search:  [Go!](#) [Browse](#) [Register a Service](#)

## New SOAP Service

1. Specify the URL to the

2. Check the details, then

After submission you will be


### Information from

Name:

WSDL Location:

Description:

Location:



This service has the

blastp (3 inputs, 1 o  
blastn (3 inputs, 1 o  
getOutput (1 input, 1  
getXML (1 input, 1 c  
runWUBlast (2 input  
checkStatus (1 input  
poll (2 inputs, 1 out  
getResults (1 input,  
getIds (1 input, 1 o  
polljob (2 inputs, 1  
doWUBlast (2 inputs



Search:

## Search Results

Search query: "blast"

### Services (2)

#### Blast [SOAP](#)

*No description*

**Provider(s):** [xml.nig.ac.jp](#)

**Submitted by:** [jits](#) (41 minutes ago)

#### Blast\_Service [SOAP](#)

Blast @ IBCP (<http://gbio-pbil.ibcp.fr>)



Search:

## Service: Blast [SOAP](#)

**Submitted by:** [jits](#) (about 1 hour ago)

**Unique code:** blast\_927483

*No description*

**Latest Version Endpoint(s):**

<http://xml.nig.ac.jp/xddb/Blast>

**All Location(s):**

Japan 

**All Provider(s):**

[xml.nig.ac.jp](#)

**WSDL Location:** <http://xml.nig.ac.jp/wsdl/Blast.wsdl>

## Operations (6):

**Name:** searchParam

SeekDa!

BioSapien    myGrid

Feed

EMBRACE

Migrate

Feed and  
Cross-link

Scrape

BioMOBY  
Central

BioLinks

DAS Registry

*BioCatalogue*   
"The Life Science Web Services Registry"

myExperiment  
Code Base



# Integration Pilots

Workflow analytics

Alternative access



REST API

Discovery access

Curation application

Service use feeds



**Taverna**  
Workflow Management System

# So why is it taking so damn long to get here?

- The final 9 yards and 80:20 rule.
- All or nothing.
- Dedicated resources and best intentions.
- Content, content, content.
- Being too damn, and unnecessarily, clever.

A social activity

# BioCatalogue Team



Hamish McWilliams



Rodrigo Lopez



Thomas Laurent



Mark Wilkinson



Carole Goble



Eric Nzuobontane



Holger Lausen



Franck Tanoh



Jiten Bhagat

# Further information

- <http://www.biocatalogue.org>
- Join our friends
- Supply technology!
- Carole Goble, Robert Stevens, Duncan Hull, Katy Wolstencroft, Rodrigo Lopez, *Data Curation + Process Curation = Data Integration + Science*, Briefings in Bioinformatics, in press