

Web Services - A Seamless Continuation of Other Technologies

Martin Senger
senger@ebi.ac.uk

Abstract

- Web Services is a technology applicable for computationally distributed problems, including access to large databases
 - What other technologies were/are available and how they compare with Web Services?
 - How to build on top of existing technologies?
- Two examples of Web Services deployed at EMBL-EBI
- The main buzzwords:
 - Integration & Standardization

What are Web Services?

It's Thursday – so you know already...

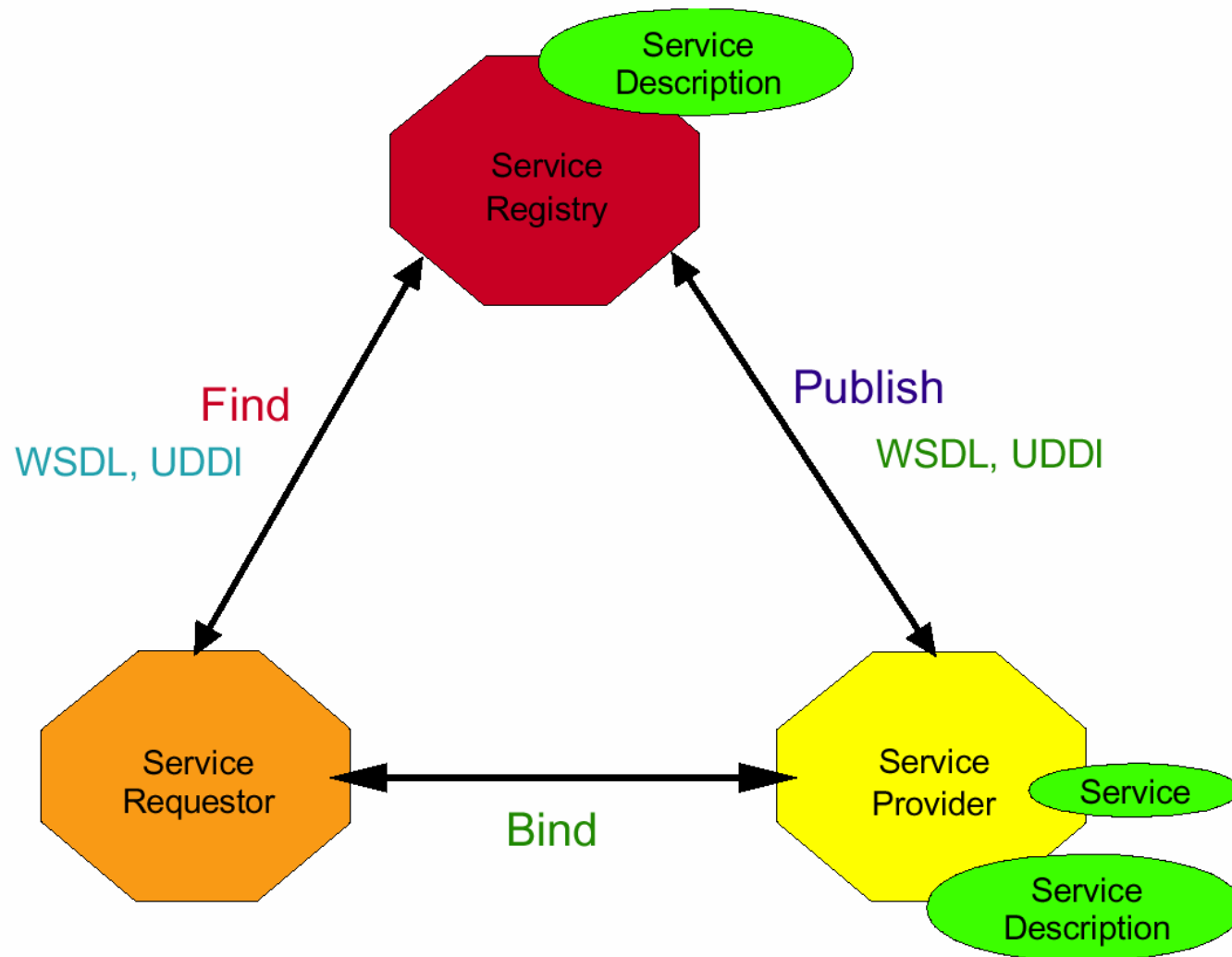
W3C (working group) definition

- *"A Web service is a software application identified by a URI, whose interfaces and bindings are capable of being defined, described and discovered as XML artefacts. A Web service supports direct interactions with other software agents using XML based messages exchanged via internet-based protocols."*
- <http://www.w3c.org/TR/2002/WD-wsa-reqs-20020819>

Dion Almaer, in an article for O'Reilly, May 2002:

- Web services have been a buzzword for a while. A friend used to say:
 - "Web services are like high school sex. Everyone is talking about doing it, but hardly anyone is, and those that are, probably aren't doing it well."
- These days, though, Web services are moving to college, so to speak, and lots of people are starting to "do it" more often and better than before. Tools are maturing, and creating and working with Web services isn't all that hard anymore.

Web Services Architecture

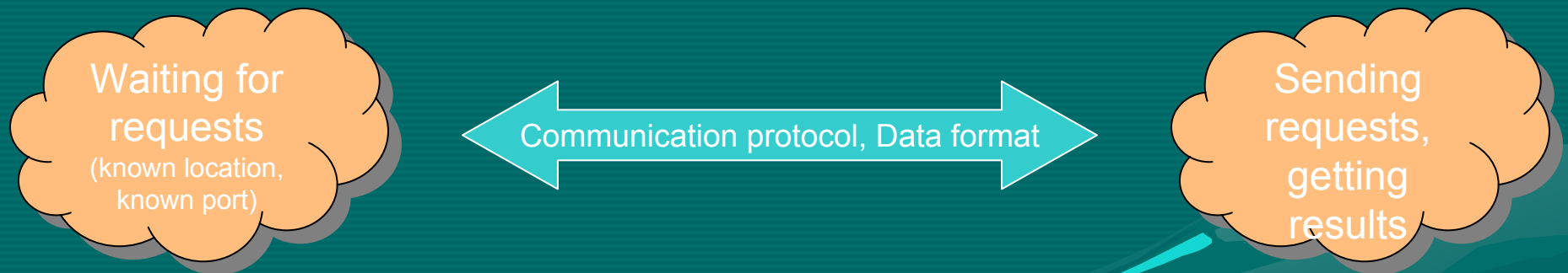


Let a program “click on a web page”

Other technologies look
similar...

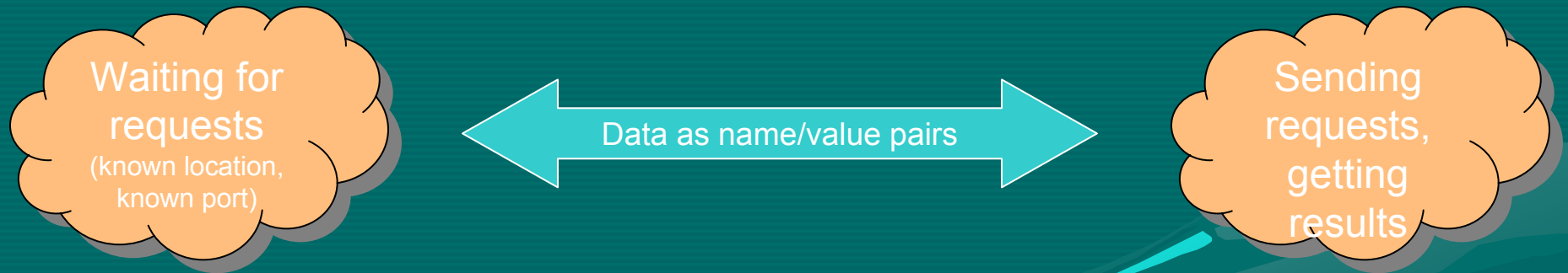
And you have already invested your
money...

Distributed architecture



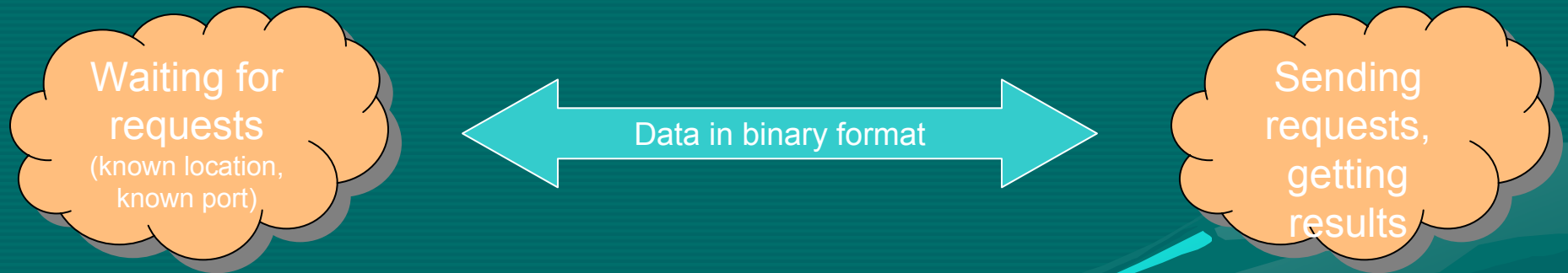
- Basic questions are:
 - Who is sending requests: a human “clicker” or a program
 - What kind of protocol to use, and what data to transmit
 - What to do with requests on the server side

Traditional CGI-based approach



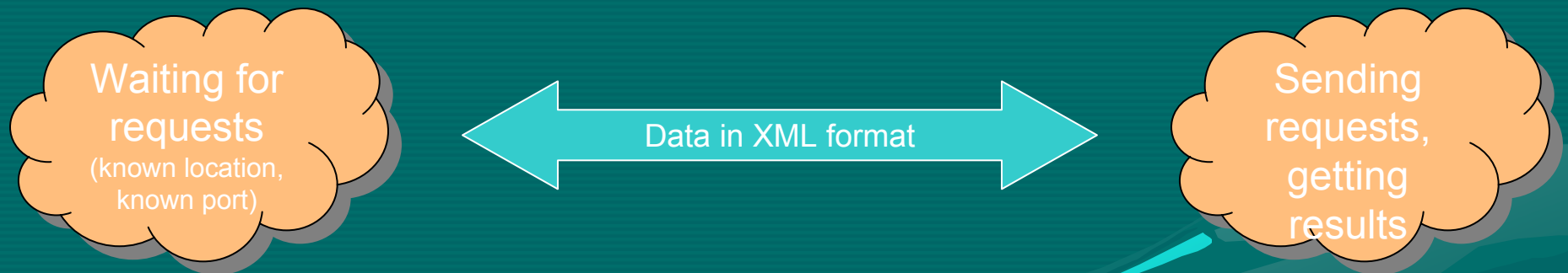
- cgi-bin scripts:
 - Data transmitted as name-value pairs (HTML forms)
 - Transport over (state-less) HTTP protocol
 - no standards for keeping user sessions (state-fullness)
 - server side: a script is called

CORBA-based approach



- CORBA:
 - Data transmitted as objects (at least it looks like that)
 - Transport (usually) over well standardised IIOP protocol
 - user sessions (state-fullness) very inter-operable
 - server side: an RPC call is made

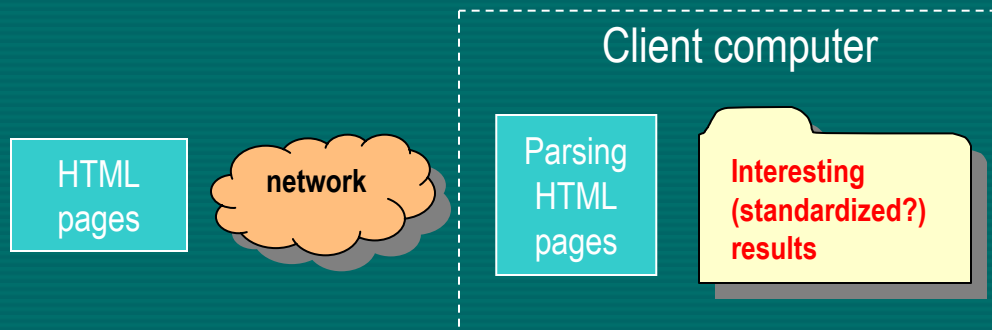
SOAP-based communication



- SOAP:
 - Data in a well-defined XML format
 - Transport over various protocols
 - HTTP, SMTP are the most used, perhaps because they are firewall-friendly
 - server side: either an RPC call or a message delivered

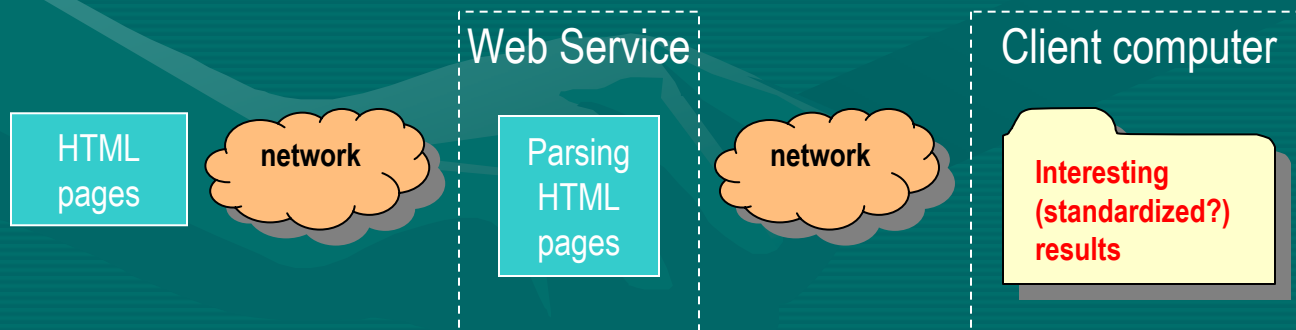
Web Services & cgi-bin scripts

Traditional screen-scraping



- Minimal network overhead
- No need to learn any new technology
- No need to define a new API

Adding a Web Service



- Much less dependent on the implementation
- Language independence for clients
- Easier integration of more resources

WS & CORBA: What is similar

- The programming effort and maintainability is roughly the same both for Web Services and CORBA
 - For CORBA I need an ORB
 - ...but do you know anybody doing WS without a SOAP toolkit?
 - For CORBA I need an IDL compiler
 - ...not always (ask Perl folks)
 - ...for WS you use frequently stubs generated from WSDL
 - ...similar answers for valuetype/custom encoding, etc.

WS & CORBA: What is *(IMHO)* better

- WS are easier to deploy because of their firewall-friendliness
- WS are quite well marketed (both from IT companies and Open Source projects)
- Integration of WS into workflows seems to be very dynamic and very real topic
 - comparing with CORBA Components (are they sleepy or I just do not know enough about them?)

WS & CORBA: What is *(IMHO)* worse

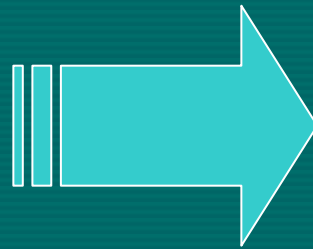
- Peer-to-peer access is problematic
 - notification by “server-push” is harder to achieve
- User sessions (server’s state-fullness) are less standardised
 - ...and therefore less inter-operable
- The “object view of world” is artificial
- Many parts yet-to-be-done (notification, transactions, security, etc.)



So what?

“Should I jump yet again into another
middleware?”

Yes: Solution A



- Throw the baby (i.e. CORBA) out with the bathwater
- ...and start from the scratch
 - not too bad if you are afraid of legacies
 - ...but may be costly

Yes: Solution B

- Combine the existing projects with a new Web Services layer; in most cases it is not so difficult
- Apply existing standards to new Web Services projects
 - think MDA – it may help, even without the whole OMG adoption process
- Let's look at some examples...

(Two) Web Services @ EMBL-EBI

European Molecular Biology Laboratory –
European Bioinformatics Institute
(Cambridge, UK)

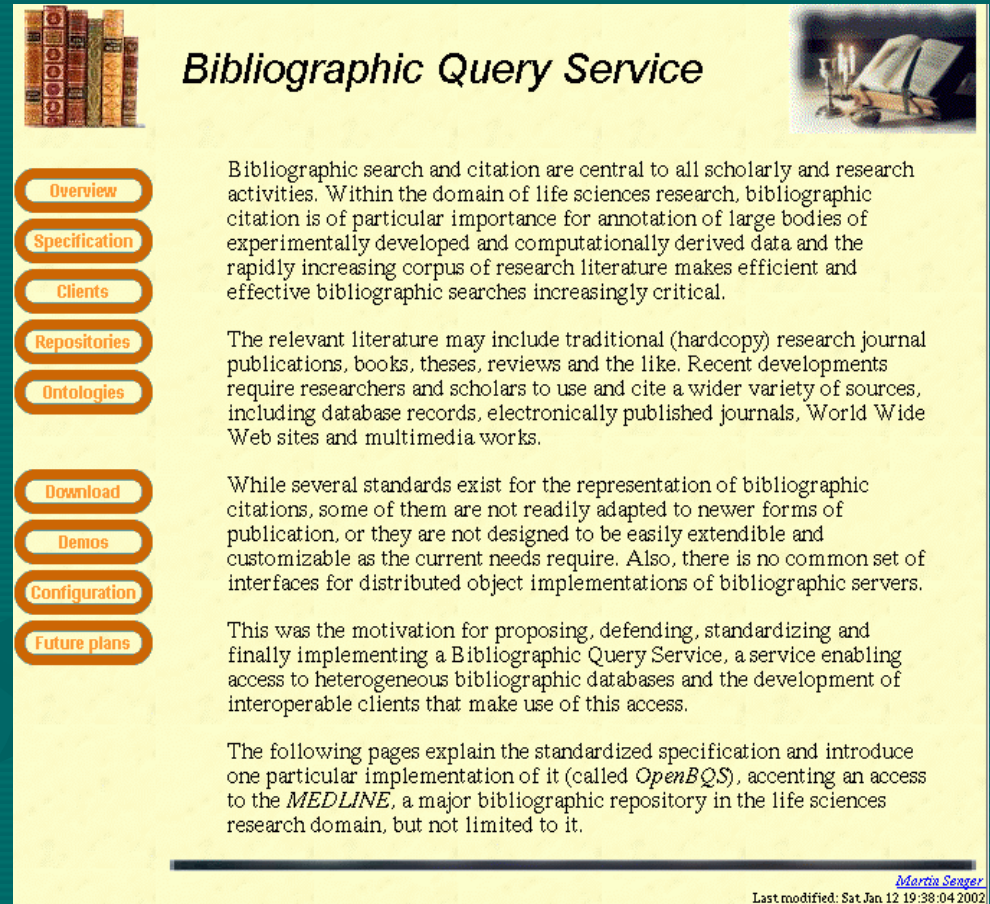
Roots, reasons...

- EBI has a long history of distributed computing
- Web Services are “in demand”
- An access both to data resources and to the analyses
 - everything is public, usually an open source
- Compositions are important
 - covering semantically rich space of the EBI activities
 - allowing seamless access to the “third-party” services

Use case 1: OpenBQS

<http://industry.ebi.ac.uk/openBQS>

- access to MEDLINE
- based on an approved standard (OMG) for accessing bibliographic data
- explores session management allowing step-by-step query refinement
- also part of the bioperl



The screenshot shows the 'Bibliographic Query Service' website. It features a navigation menu on the left with buttons for Overview, Specification, Clients, Repositories, Ontologies, Download, Demos, Configuration, and Future plans. The main content area includes an introduction to the service, a list of relevant literature types, a discussion on standards for bibliographic citations, and a description of the service's motivation and implementation. The website also includes a footer with the author's name, 'Martia Senger', and the date 'Last modified: Sat, Jan 12 19:38:04 2002'.

Bibliographic Query Service

Bibliographic search and citation are central to all scholarly and research activities. Within the domain of life sciences research, bibliographic citation is of particular importance for annotation of large bodies of experimentally developed and computationally derived data and the rapidly increasing corpus of research literature makes efficient and effective bibliographic searches increasingly critical.

The relevant literature may include traditional (hardcopy) research journal publications, books, theses, reviews and the like. Recent developments require researchers and scholars to use and cite a wider variety of sources, including database records, electronically published journals, World Wide Web sites and multimedia works.

While several standards exist for the representation of bibliographic citations, some of them are not readily adapted to newer forms of publication, or they are not designed to be easily extendible and customizable as the current needs require. Also, there is no common set of interfaces for distributed object implementations of bibliographic servers.

This was the motivation for proposing, defending, standardizing and finally implementing a Bibliographic Query Service, a service enabling access to heterogeneous bibliographic databases and the development of interoperable clients that make use of this access.

The following pages explain the standardized specification and introduce one particular implementation of it (called *OpenBQS*), accenting an access to the *MEDLINE*, a major bibliographic repository in the life sciences research domain, but not limited to it.

Martia Senger
Last modified: Sat, Jan 12 19:38:04 2002

CORBA => WS: Object references

CORBA: a real world of objects

WS: Object references substituted by (string) identifiers

BibRefCollection

```
BibRefCollection find (String[] keywords);  
String export_all_elements ();
```

```
String find (String[] keywords);
```



```
String find (String collectionId, String[] keywords);
```

```
String export_all_elements (String collectionId);
```

Other ways for object references

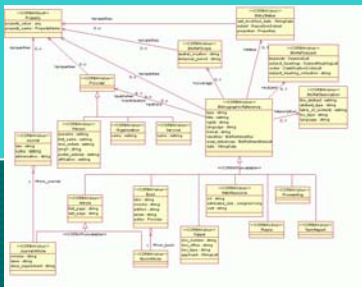
- Each object/class is represented by a separate web service with its own endpoint (URI)
 - it is the closest to the original CORBA API
 - but it must deal with the life-cycle of web services (deployment/undeployment) which will be very little portable (strong dependence on the used toolkit)
 - and it must consider how “private” the new service instance should be
- Re-design your API in order to avoid returning object references at all

CORBA => WS: Complex data structures

- WS allow complex data structures
 - but it may require to write specific data encoders which puts burden on clients you do not control
- Better to use common data types and test language interoperability

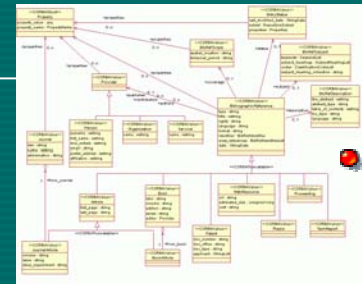
BibRefCollection

BibliographicReference **get_by_id** (String id);



String **get_by_id** (String id);

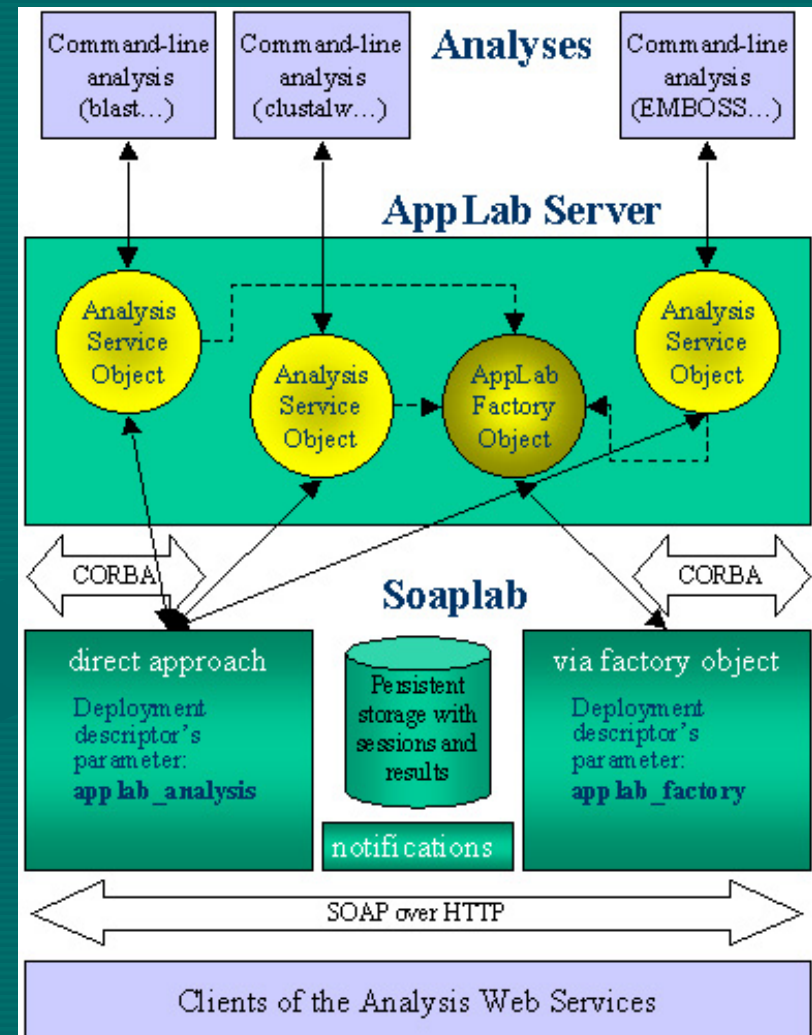
it's the client task to parse the returned XML representation of a bibliographic citation



Use case 2: Soaplab

<http://industry.ebi.ac.uk/soaplab>

- access to analysis tools
- based on an approved standard for accessing remote applications
- access to EMBOSS programs, to NCBI-Blast and WU-Blast services
- an example of technology collaboration (CORBA - Web Services)



CORBA => WS: Callbacks

- Drastic changes in API needed if “server-push” notification is required because:
 - the WS client uses different libraries than the server
 - there is no common and standardized ORB used
 - there is (usually) no servlet engine running on the client side
 - the WS client may be behind its own firewall
 - each client may prefer a different way how to be contacted
- “A” solution:
 - add a protocol negotiation into the WS API
 - be prepared to use also the SMTP protocol (firewall-friendlier)

Conclusions

- Distributed computing is inevitable
- More accesses by programs than by clicking on hyperlinks
- More technologies of distributed architecture will collaborate
- The better standards we have the better results we'll get
- Web Services is not a new hype but a trend to follow

Acknowledgement

- EMBL-EBI e-Science group (Alan Robinson, Tom Oinn, David Starks-Browning)
- EMBL-EBI Industry Support Group (Lichun Wang)
- Support and funding from the multi-organisational project “myGrid” (<http://www.mygrid.info>) funded by the EPSRC (<http://www.epsrc.ac.uk>) as part of the UK’s Research Councils (<http://www.research-councils.ac.uk>)