WEB SERVICES EN BIOINFORMÁTICA
DAS, BIOMOBY Y TAVERNA

TEORÍA

José Manuel Rodríguez Carrasco
José María Fernández González
Biología Estructural y Bioinformática, CNIO
GN2 (CNIO), INB

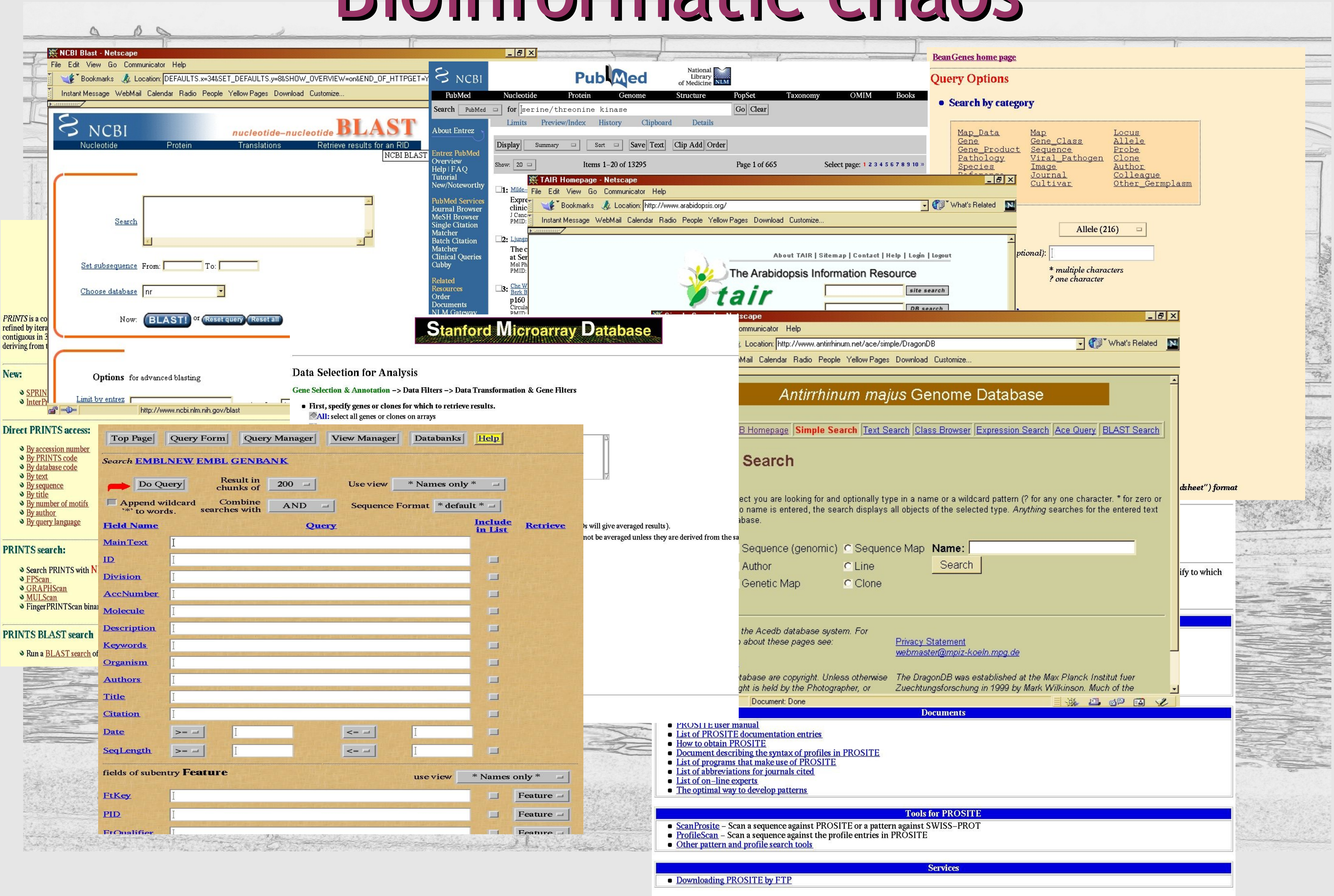
cnio
Centro Nacional
de Investigaciones
Oncológicas

VNIVERSITAS COMPLVTENSIS MATRITENSIS

INB

José Manuel Rodríguez Carrasco (jmrodriguez@cnio.es)
José María Fernández González (jmfernandez@cnio.es)

Bioinformatic chaos

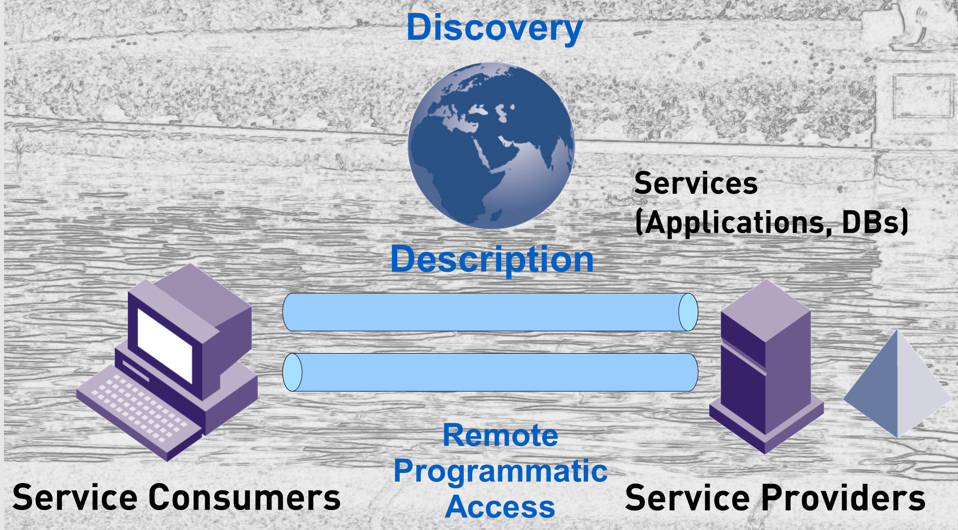


Current (& past) Practices

Do you have data?

Do you have tools?

Publish a web page



Discovery

Services (Applications, DBs)

Description

Remote Programmatic Access

Service Consumers

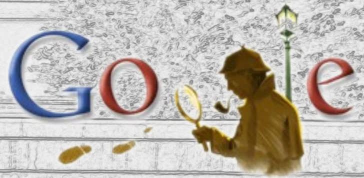
Service Providers

Bioinformatics integration: State of the Art

A Web Page is the *de facto* standard

Discovery:

- Word of mouth
- Web directories
- Web searches
- Paper publications



Description:

- Word of mouth
- Web documentation / examples / tutorials / courses
- Paper publications

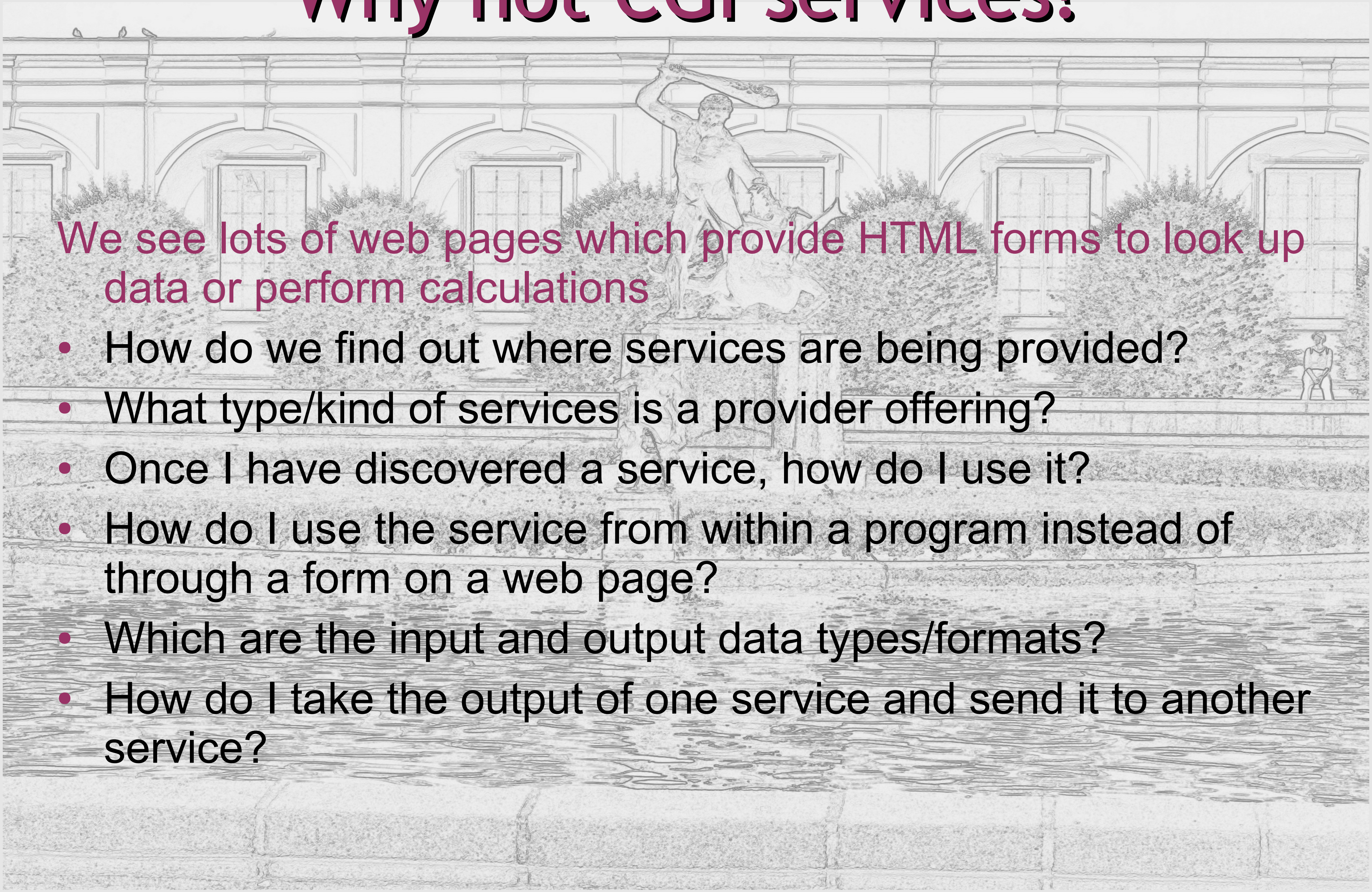
Data Transfer and Message Format:

- Cut & Paste© + Data reformatting

Automation:

- CGI & *Bespoke code* (ad hoc)
- APIs (usually only big Bioinformatics Projects/Institutes)

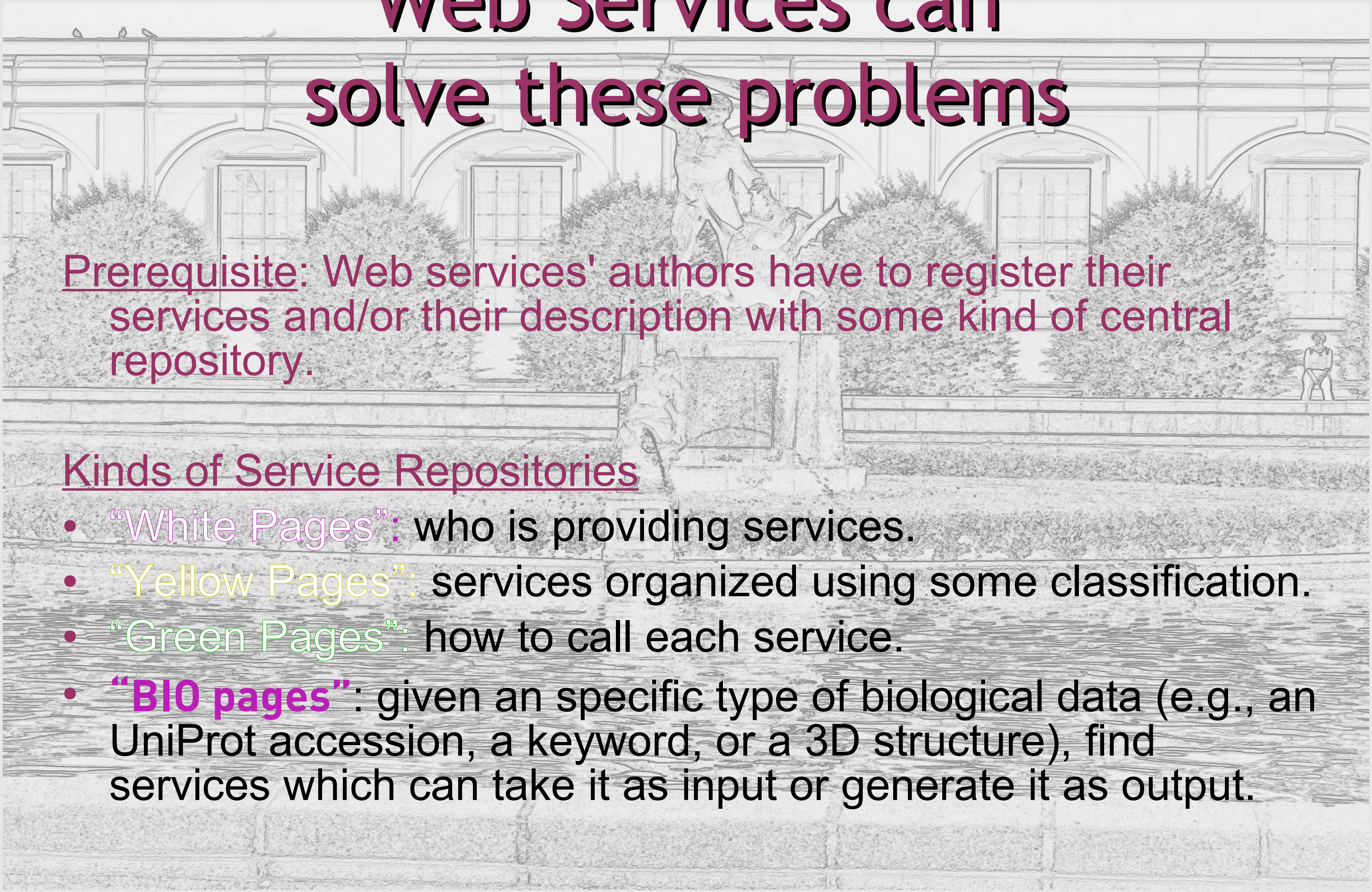
Why not CGI services?



We see lots of web pages which provide HTML forms to look up data or perform calculations

- How do we find out where services are being provided?
- What type/kind of services is a provider offering?
- Once I have discovered a service, how do I use it?
- How do I use the service from within a program instead of through a form on a web page?
- Which are the input and output data types/formats?
- How do I take the output of one service and send it to another service?

Web Services can solve these problems



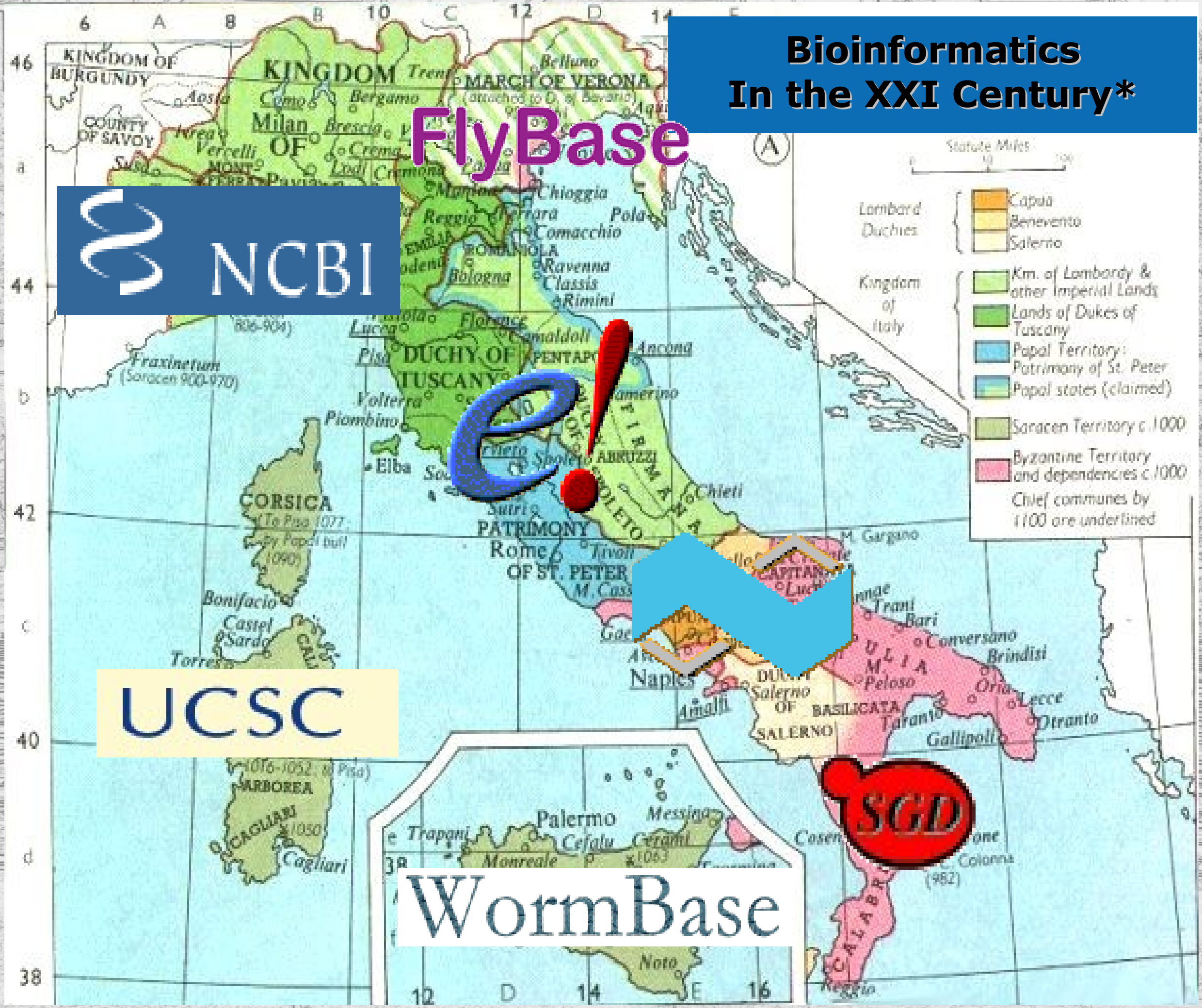
Prerequisite: Web services' authors have to register their services and/or their description with some kind of central repository.

Kinds of Service Repositories

- "White Pages": who is providing services.
- "Yellow Pages": services organized using some classification.
- "Green Pages": how to call each service.
- **"BIO pages"**: given an specific type of biological data (e.g., an UniProt accession, a keyword, or a 3D structure), find services which can take it as input or generate it as output.

The Bioinformatics Curse

(and a way to delay it)



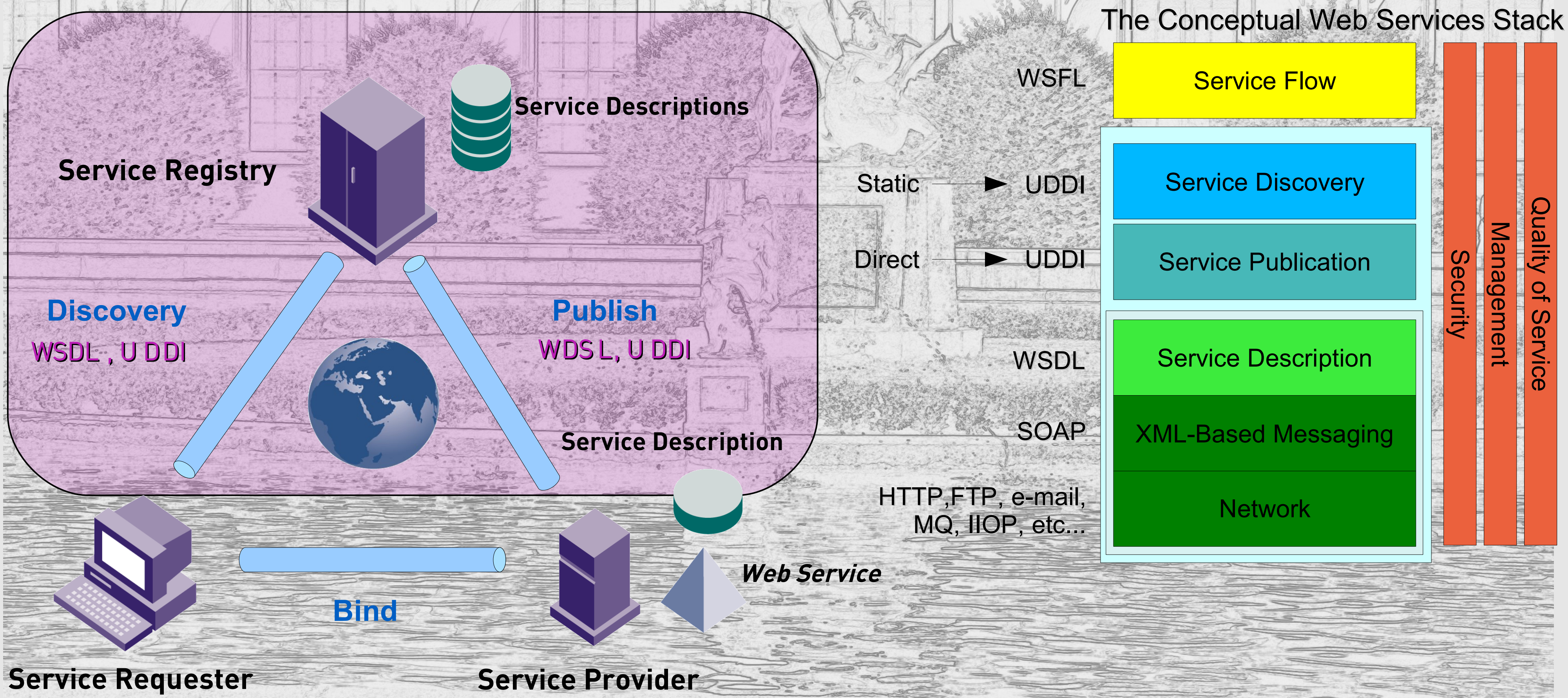
Bioinformatics In the XXI Century*

Code of Conduct for Bioinformatics

1. A Web page is an interface: Facilitate access of Web pages to scripts.
2. An interface is a contract: Give advance warning of interface changes to minimize inconvenience; use new URLs and provide temporary legacy interface.
3. Choice is good: SOAP/XML is "sweet!" — HTML is discouraged.
4. Allow batch downloads: Break data sets into bite-sized pieces to facilitate downloading.
5. Use existing file formats: Do not reinvent the wheel when multiple workable formats already exist.
6. Design sensible formats: Use common sense and simplicity.
7. Allow ad hoc queries: Make copies of databases available for direct access using native query language.

***L. Stein, "Creating a bioinformatics nation", Nature 317:119-120 (2002)**

Web Services model



Service Registry

Service Descriptions

Discovery

WSDL , U DDI

Publish

WDS L, U DDI

Service Description

Bind

Web Service

Service Requester

Service Provider

The Conceptual Web Services Stack

WSFL — Service Flow

Static → UDDI — Service Discovery

Direct → UDDI — Service Publication

WSDL — Service Description

SOAP — XML-Based Messaging

HTTP,FTP, e-mail, MQ, IIOP, etc... — Network

Security

Management

Quality of Service

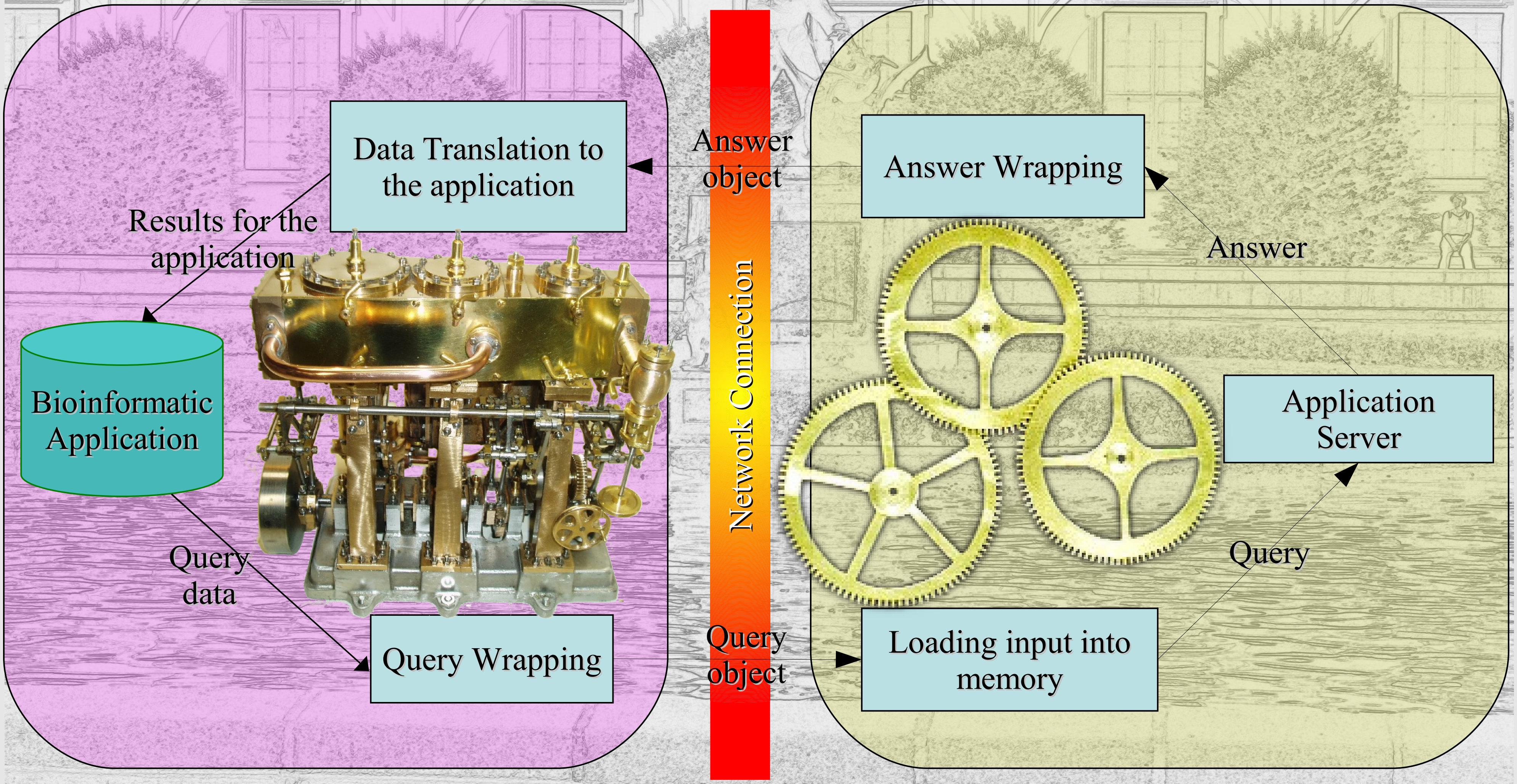
A web service is an interface which describes a set of operations which are network available using standardized XML messages*

*Web Services Conceptual Architecture, Heather Kreger, IBM Software Group, 2001

Web Services Status

- Simples, Open, Broad Industry Support
- Futuristic model, although lots of investment has already been made i.e.. Microsoft .NET
- Technical infrastructure available (assets): Existing foundation using XML-based technologies
  - **XML (*eXtensible Markup Language*):** Universal Data Transport format
  - **SOAP (*Simple Object Access Protocol*), XML-RPC:** Wire protocol, service invocation and interactions.
  - **WSDL (*Web Services Description Language*), DAML-S:** Formal service descriptions.
  - **UDDI (*Universal Description, Discovery and Integration*):** Service registry (publication) and discovery (find).
  - **XSD, XSLT (*Extensible Stylesheet Language Transformations*):** Data Types, Description and Conversion.
  - **HTML, SMTP, FTP:** Ubiquitous Communications Protocol
- **Examples**:
  - Google Web APIs (http://www.google.com/apis/ and Gmail)
  - XEMBL (http://www.ebi.ac.uk/xembl/)
  - NCBI Entrez SOAP APIs http://eutils.ncbi.nlm.nih.gov/entrez/query/static/esoap_help.html

Web Service query processing



Bioinformatic Application

Query data

Query Wrapping

Query object

Loading input into memory

Query

Application Server

Answer

Answer Wrapping

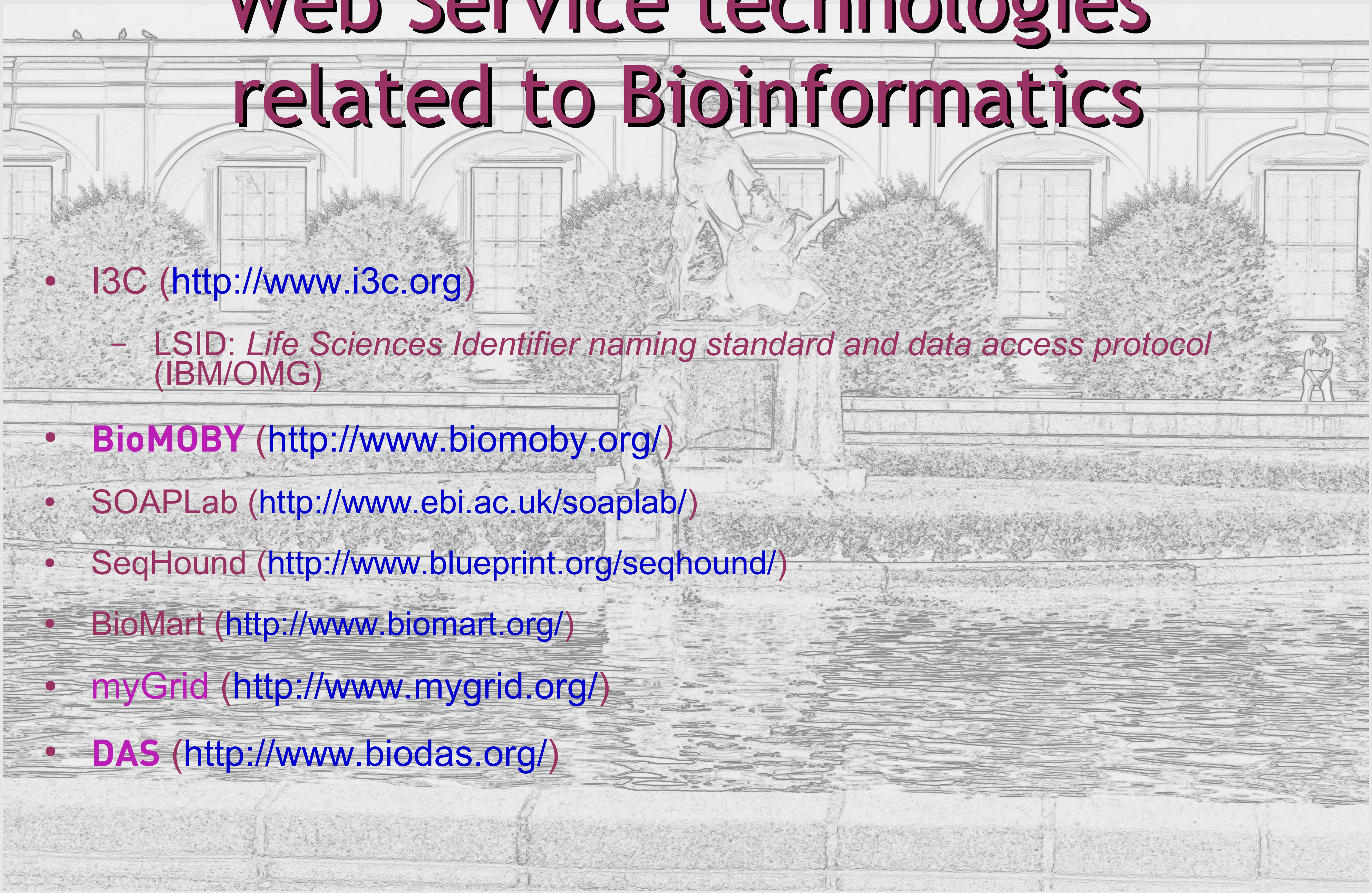
Answer object

Data Translation to the application

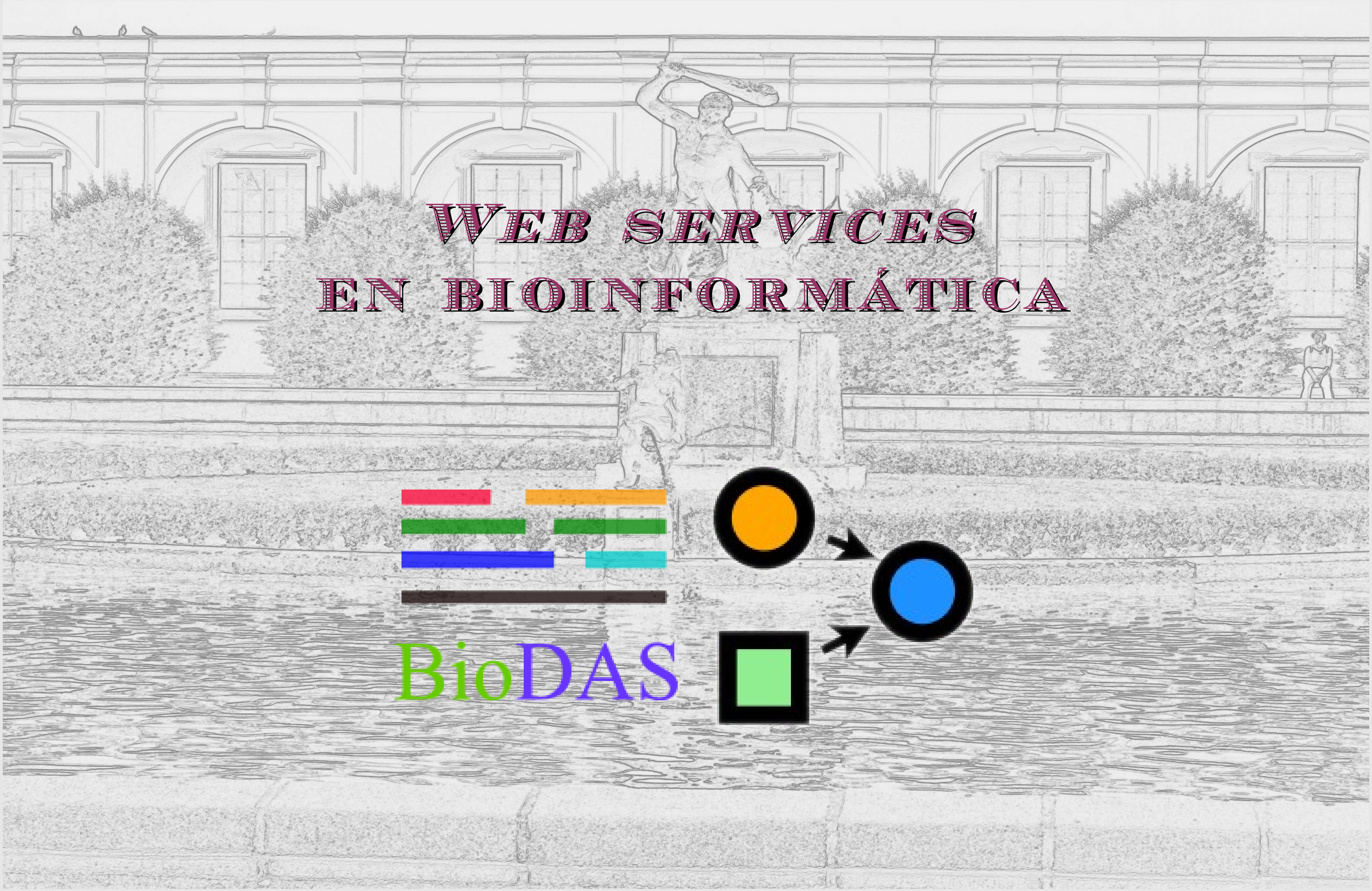
Results for the application

Network Connection

Web Service technologies related to Bioinformatics



- I3C (http://www.i3c.org)
  - LSID: *Life Sciences Identifier naming standard and data access protocol* (IBM/OMG)
- **BioMOBY** (http://www.biomoby.org/)
- SOAPLab (http://www.ebi.ac.uk/soaplab/)
- SeqHound (http://www.blueprint.org/seqhound/)
- BioMart (http://www.biomart.org/)
- myGrid (http://www.mygrid.org/)
- **DAS** (http://www.biodas.org/)



WEB SERVICES EN BIOINFORMÁTICA

BioDAS

José Manuel Rodríguez Carrasco (jmrodriguez@cnio.es)
José María Fernández González (jmfernandez@cnio.es)

DAS
(*Distributed Annotation System*)

El sistema DAS de anotación distribuida es un sistema cliente-servidor en el que cualquier cliente puede integrar información de diferentes servidores. Fue diseñado con las siguientes premisas:

- Independencia de esquema o tecnología de BD
- Independencia de tecnologías de cliente
- Servidores de referencia y de anotación no relacionados
- Gestión de inestabilidad en ensamblajes genómicos
- Sencillo de implementar

Como consecuencia de ello, tiene las siguientes características:

- Cualquier cliente DAS puede integrar información de varios sitios web remotos y mostrarla en una vista única al usuario.
- Los distintos proveedores de información necesitan coordinarse entre sí poco o nada.

La especificación 1.0 fue escrita por Lincoln Stein y Robin Dowell

¿Quién está usando DAS?

- Ensembl (muchos organismos, pero sobre todo *H. Sapiens*)
- WormBase (*C. elegans*)
- Washington University @ St. Louis (*C. elegans*)
- Flybase (*Drosophila*)
- UCSC Genome Center (*H. Sapiens*)
- TIGR (*C. elegans, H. Sapiens*)
- Cambridge University Genetics Department (*C. elegans, H. Sapiens, Drosophila*)
- DOE Joint Genome Institute (*H. Sapiens, M. Musculus*)
- NCBI (*in progress*)
- INB (*also in progress*)

José Manuel Rodríguez Carrasco (jmrodriguez@cnio.es)
José María Fernández González (jmfernandez@cnio.es)

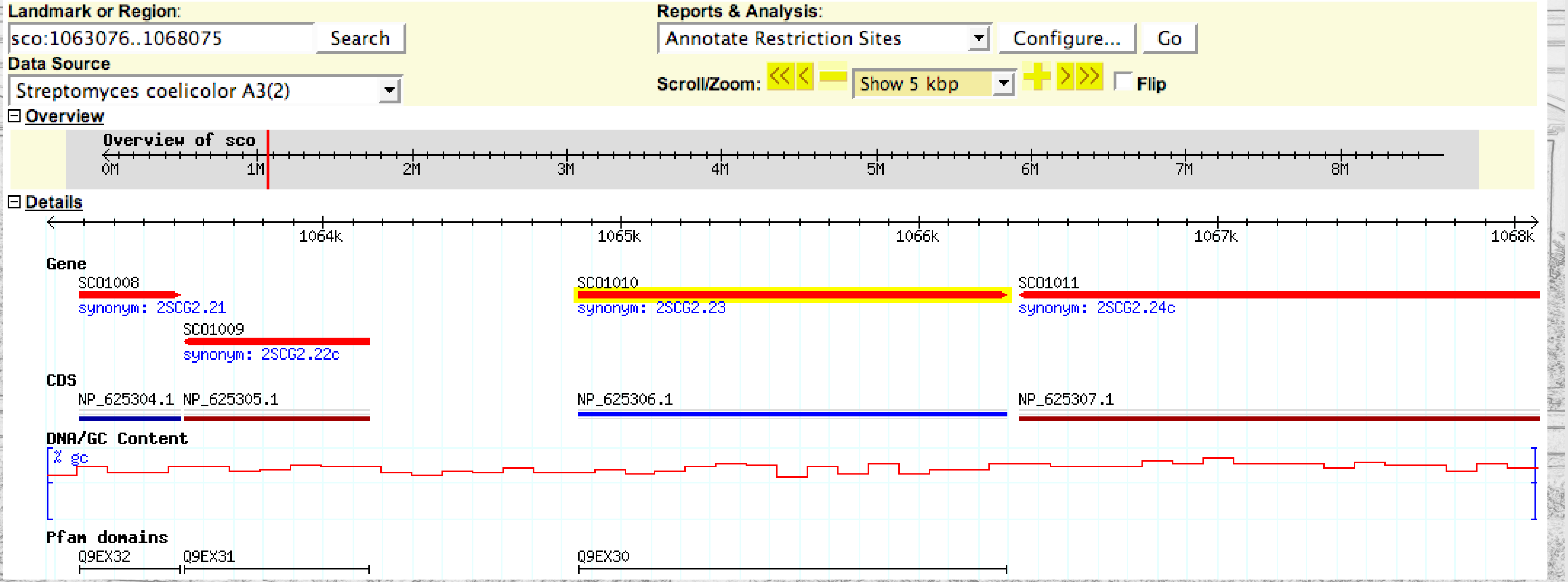
¿Qué es una anotación?

Cualquier información definible entre dos coordenadas genómicas

Genes
Exons
CDS
Restriction Sites
Ribosome binding sites
Repeats
PolyA signals
…….
Scientific references (papers)

- Sequencing Process Annotations
    - Posiciones de clones, secuencias de fin de BAC
- Análisis *In Silico*
    - Predicción de genes, Homologías
- Datos Experimentales
    - Knockouts, resultados de Microarray
- Otros
    - Grupos de Discusión, referencias PubMed, GO Terms

¿Cómo se anota?



Landmark or Region: sco:1063076..1068075

Data Source: Streptomyces coelicolor A3(2)

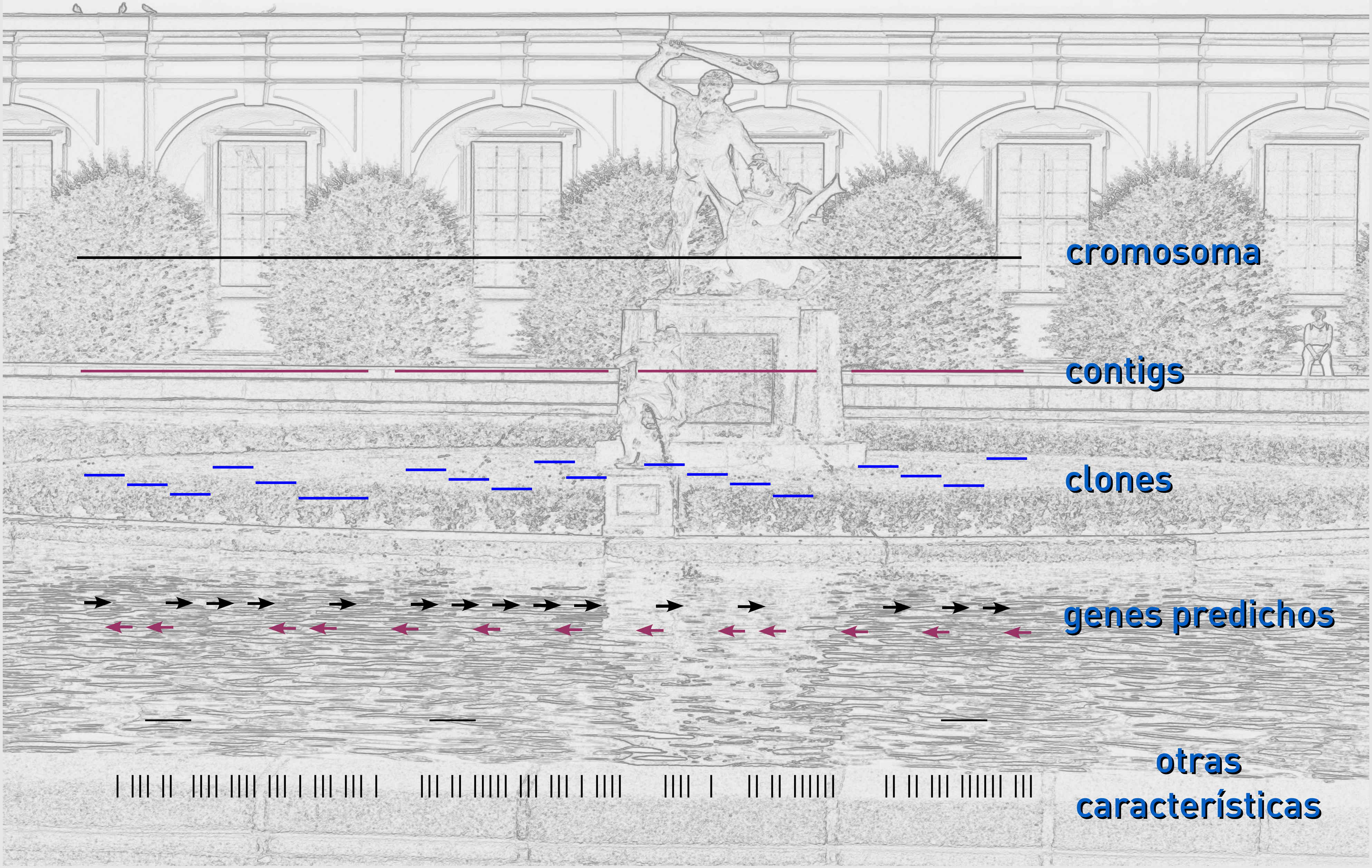
Reports & Analysis: Annotate Restriction Sites

Scroll/Zoom: Show 5 kbp

☐ Flip

ID:	gene:SC01009
Synonym:	2SCG2.22c
Start, Stop:	1063520, 1064140
Orientation:	-
Category:	Transcription
Type:	Gene
Method:	GeneFinder v1.3
Group:	2SCG2

Muchos sistemas de coordenadas



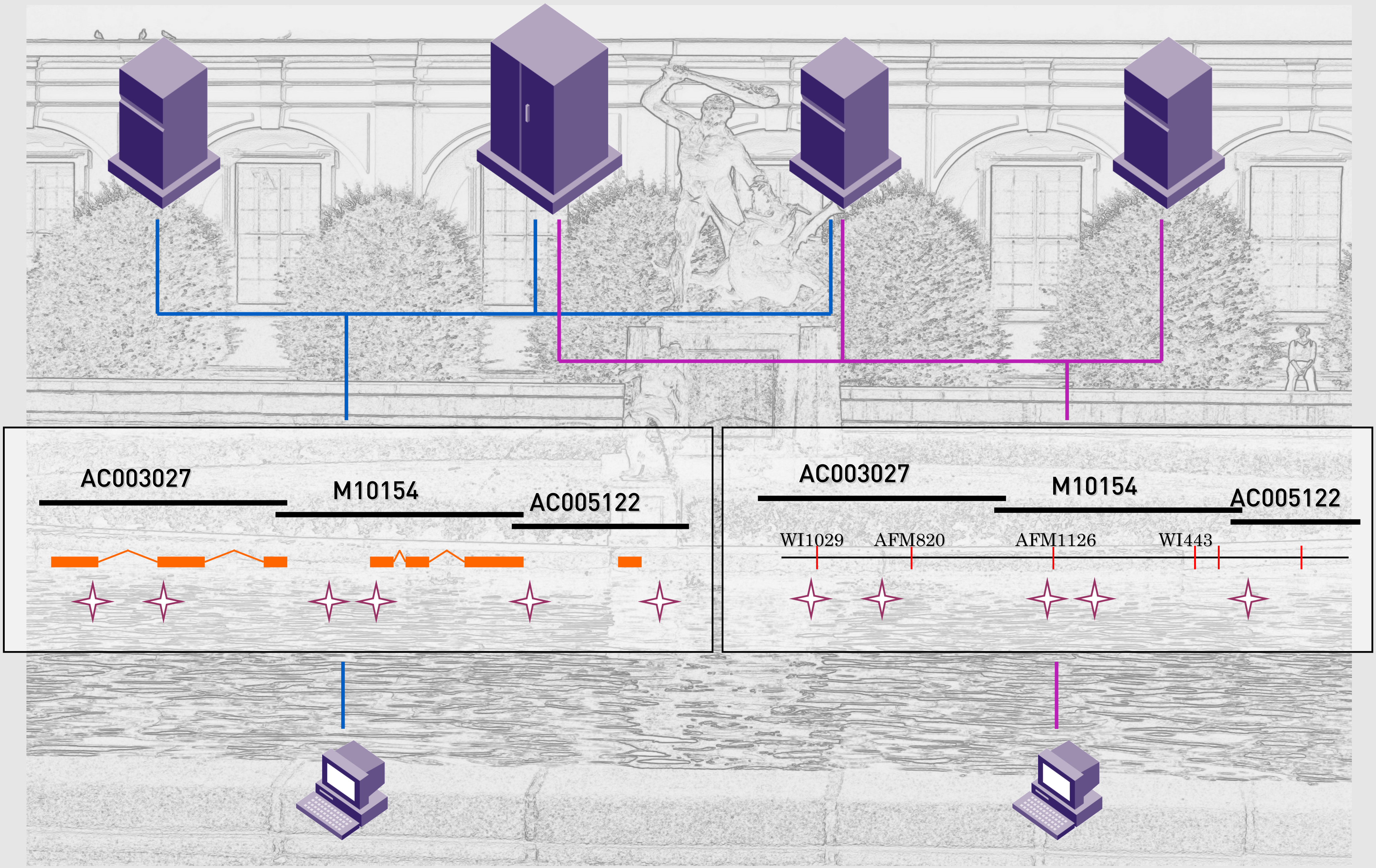
cromosoma

contigs

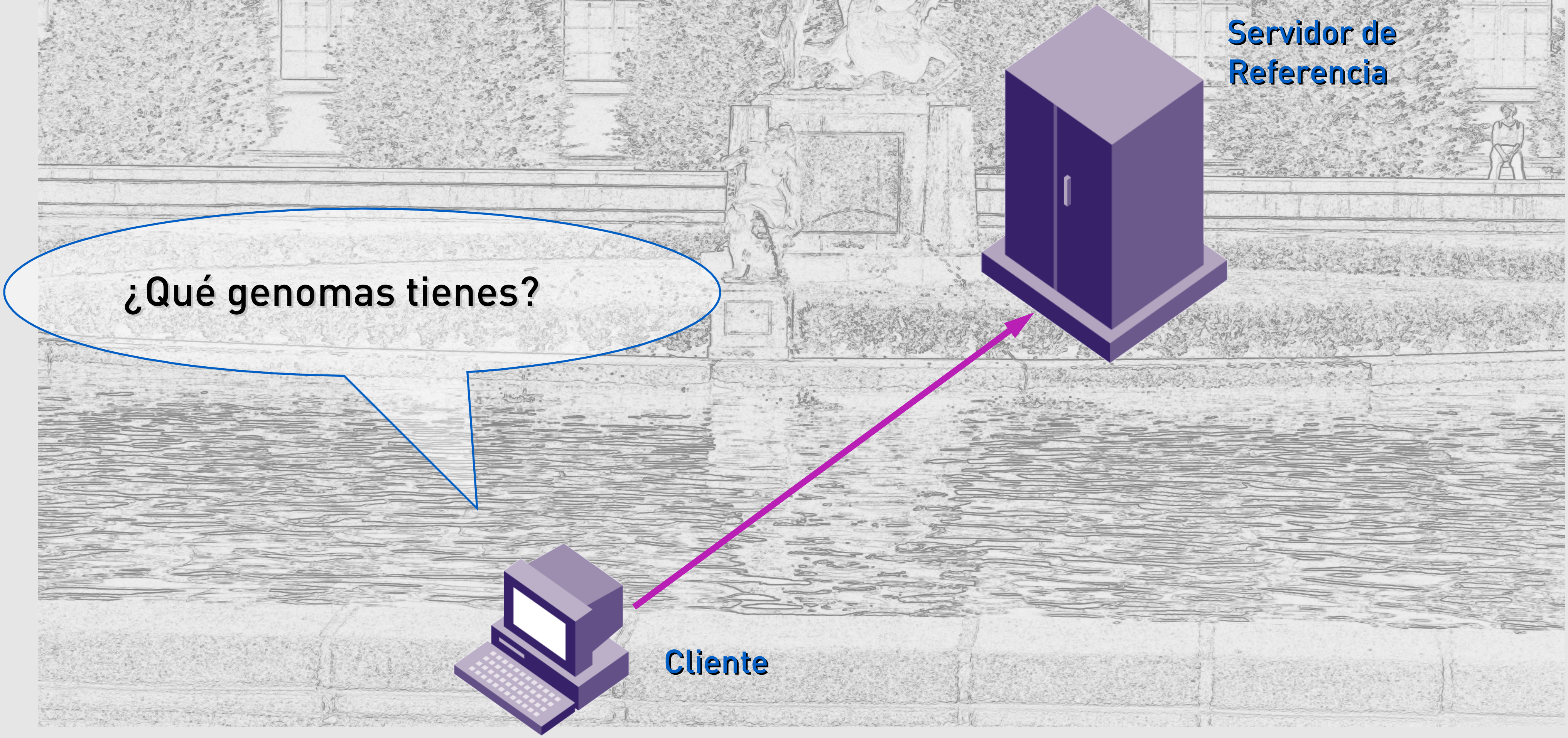
clones

genes predichos

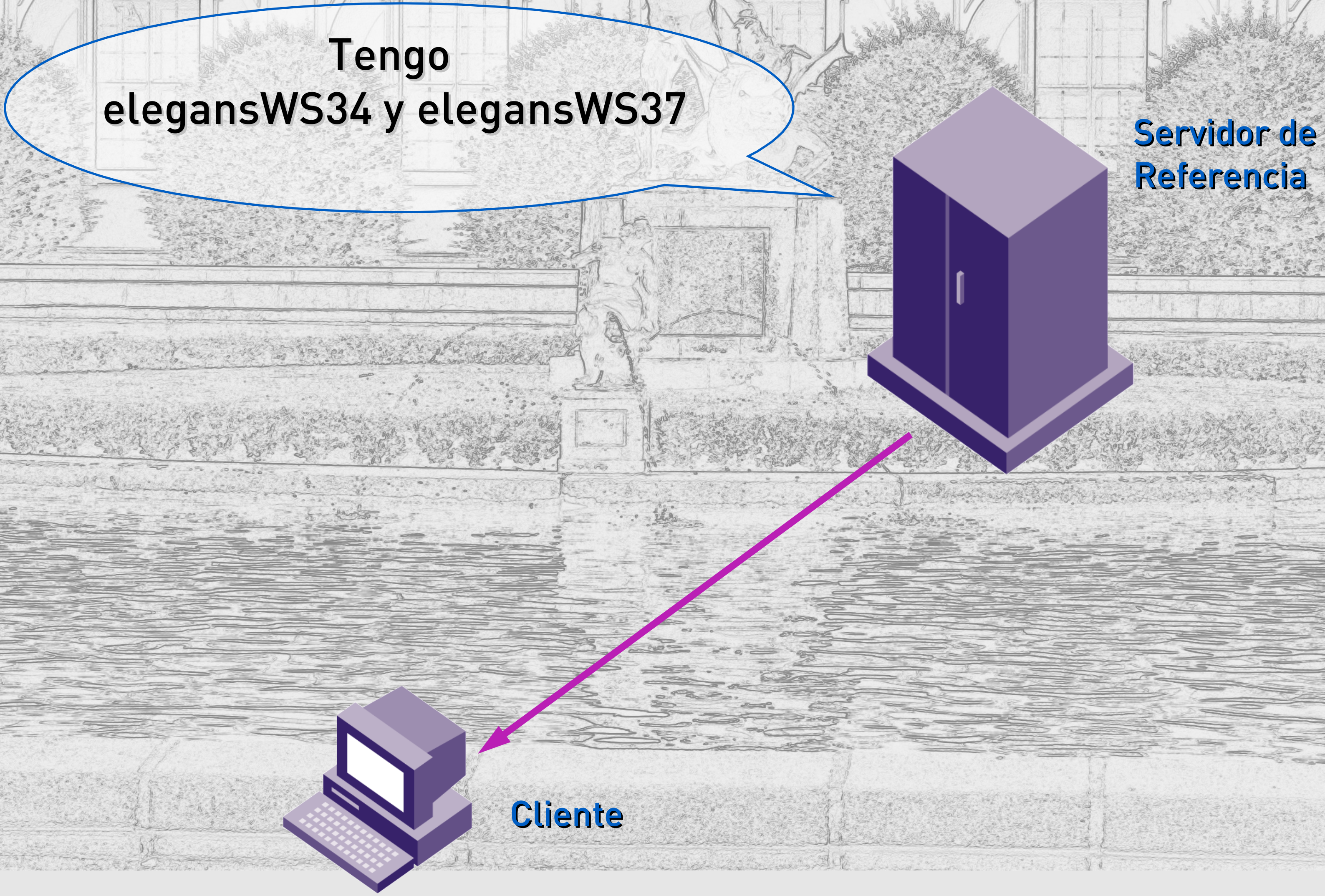
otras características


AC003027
M10154
AC005122
AC003027
M10154
AC005122
WI1029
AFM820
AFM1126
WI443

Protocolo DAS
(1) *Data Sources*



Protocolo DAS
(2) Respuesta *Data Sources*

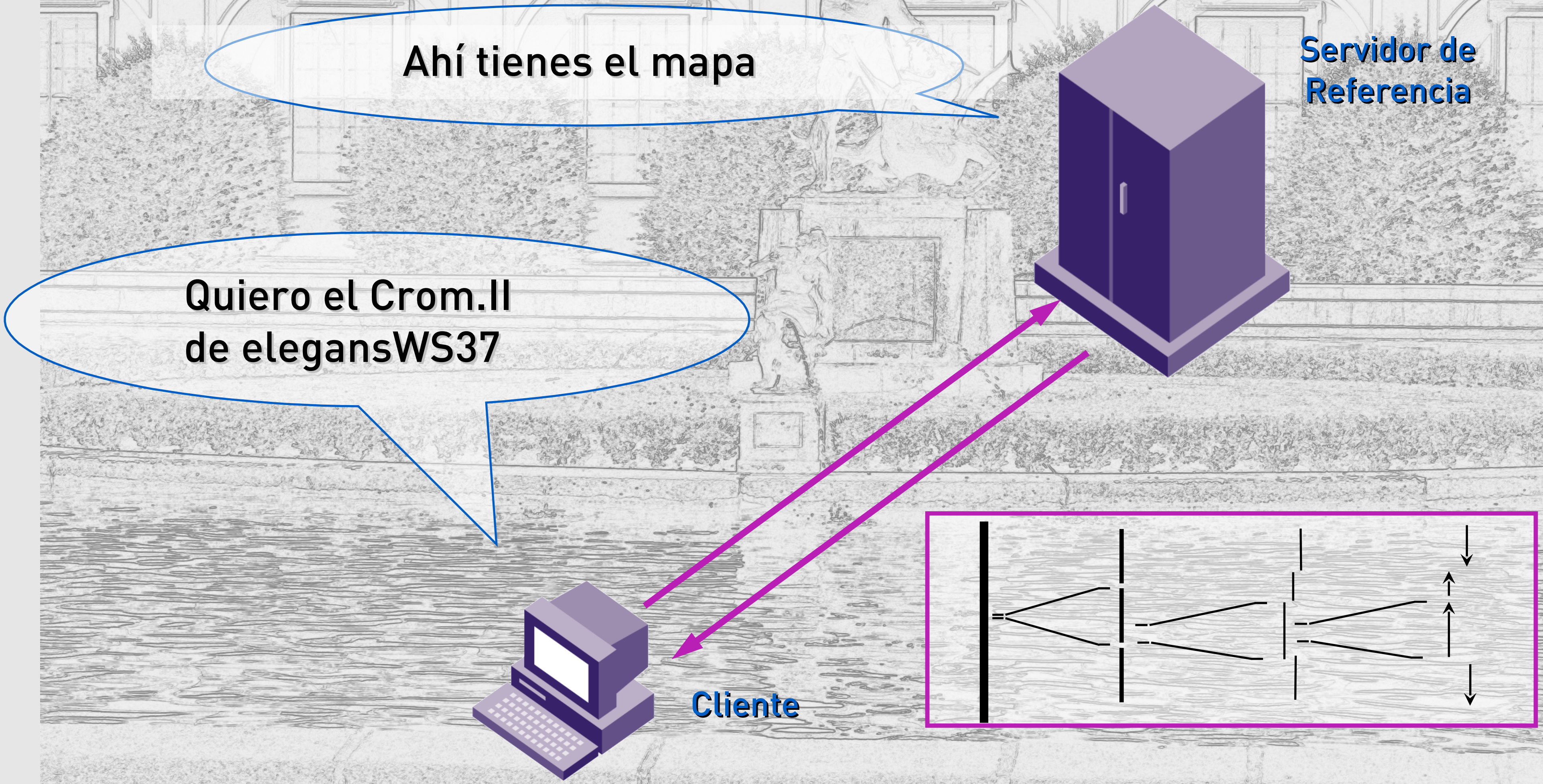


Tengo
elegansWS34 y elegansWS37

Servidor de Referencia

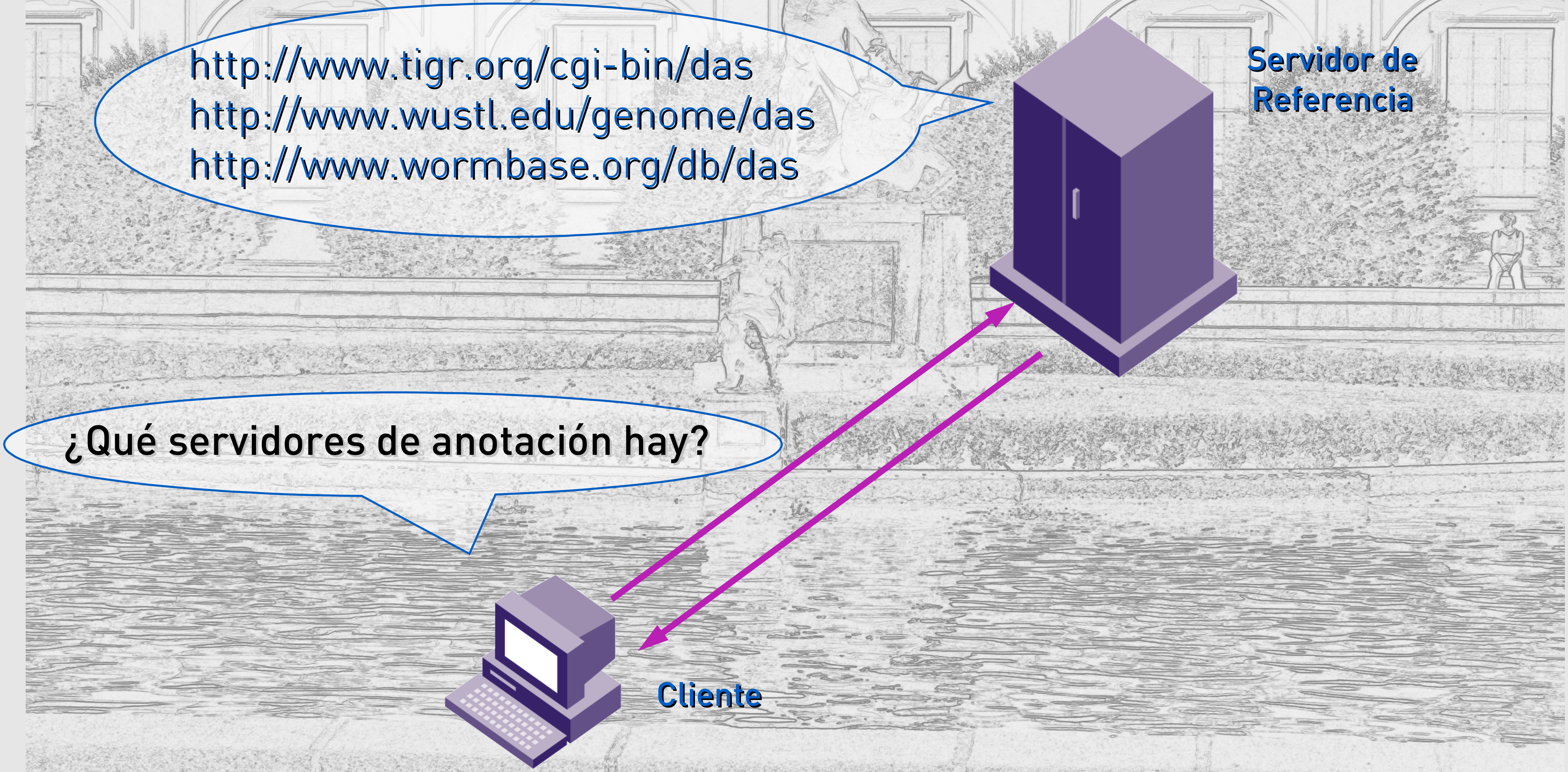
Cliente

Protocolo DAS
(3) Recuperar mapa de anotaciones



José Manuel Rodríguez Carrasco (jmrodriguez@cnio.es)
José María Fernández González (jmfernandez@cnio.es)

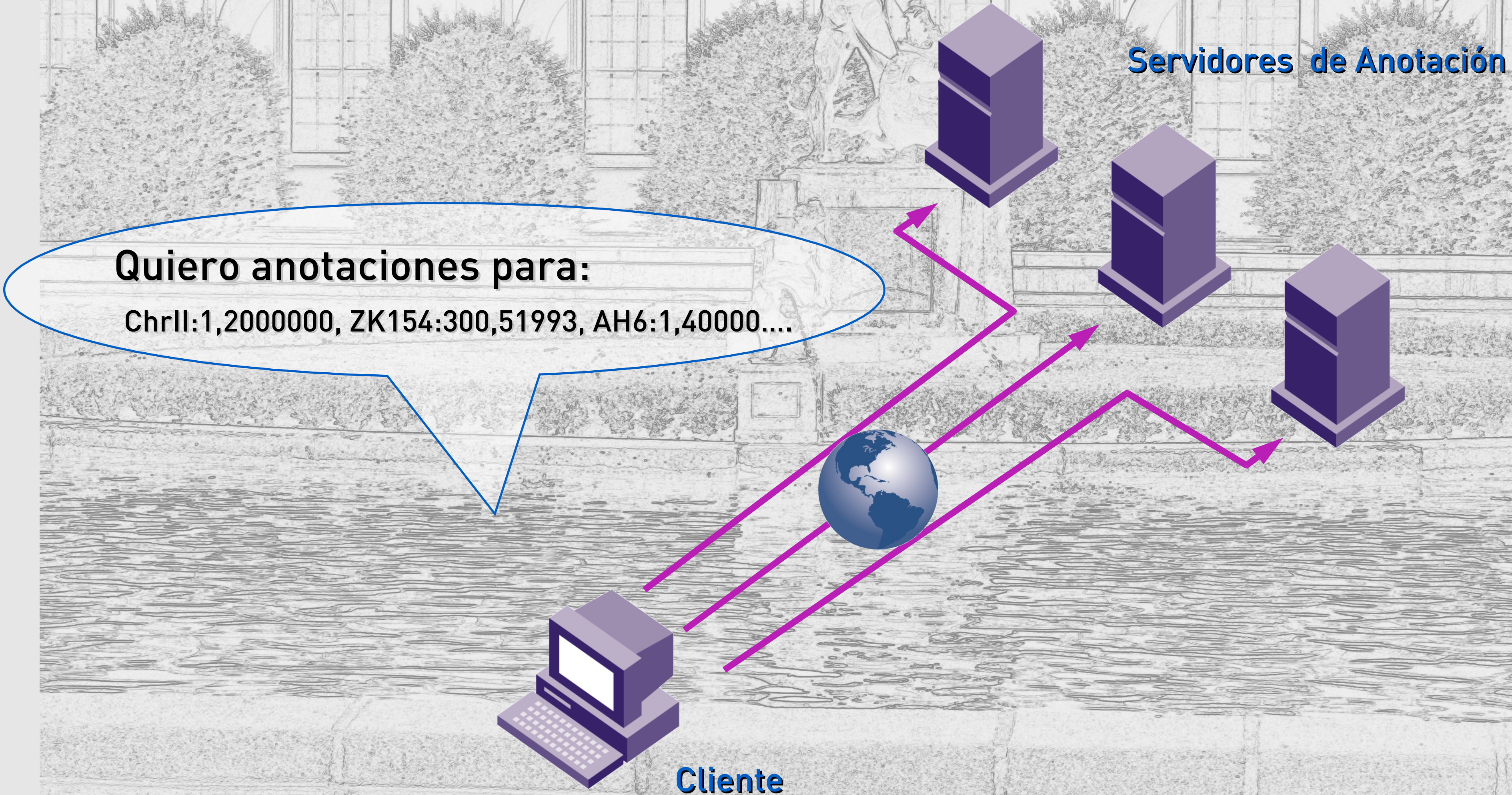
Protocolo DAS (4) Servidores de anotaciones



http://www.tigr.org/cgi-bin/das
http://www.wustl.edu/genome/das
http://www.wormbase.org/db/das

¿Qué servidores de anotación hay?

Protocolo DAS
(5) Solicitar anotaciones

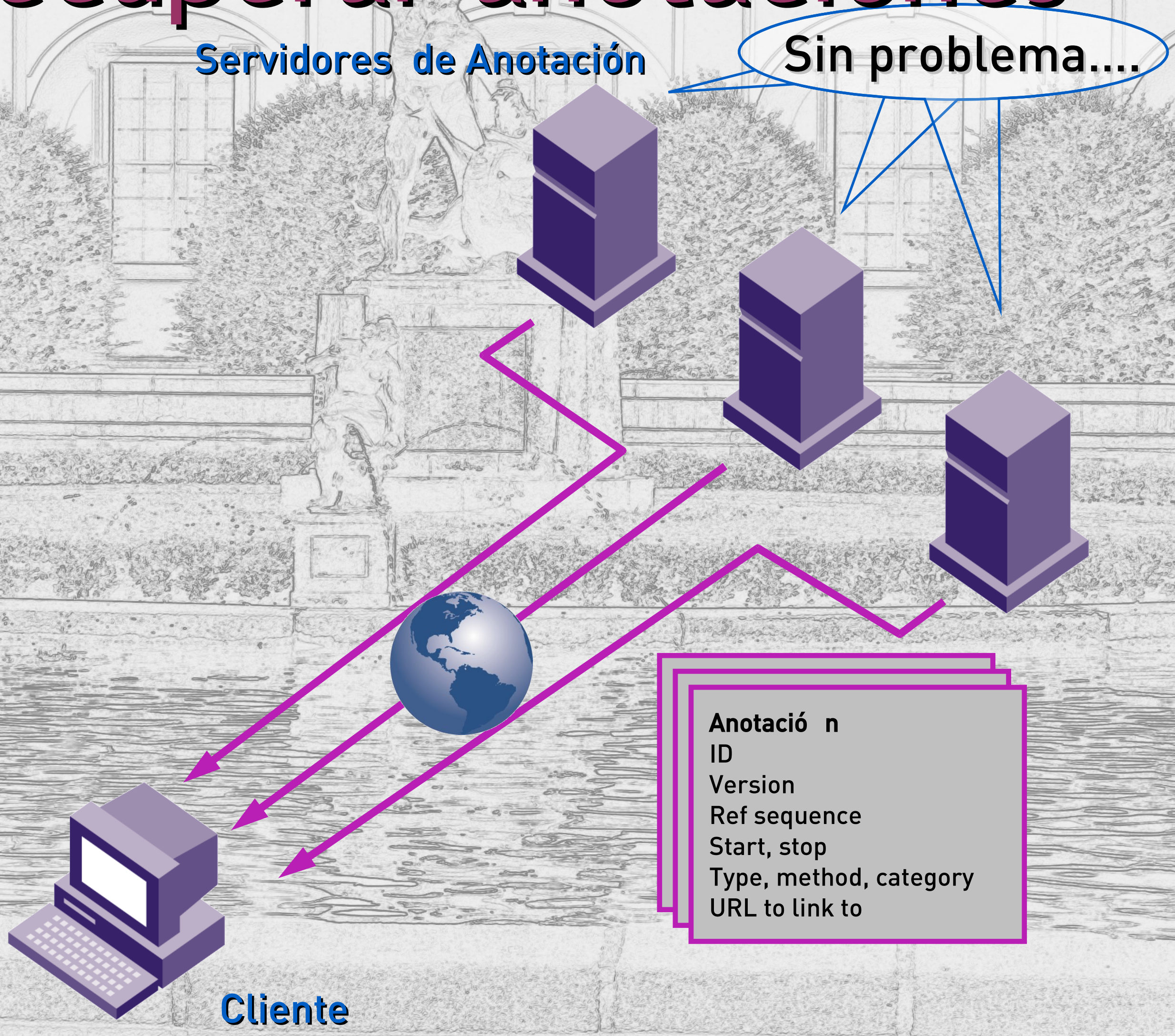


Quiero anotaciones para:

ChrII:1,2000000, ZK154:300,51993, AH6:1,40000....

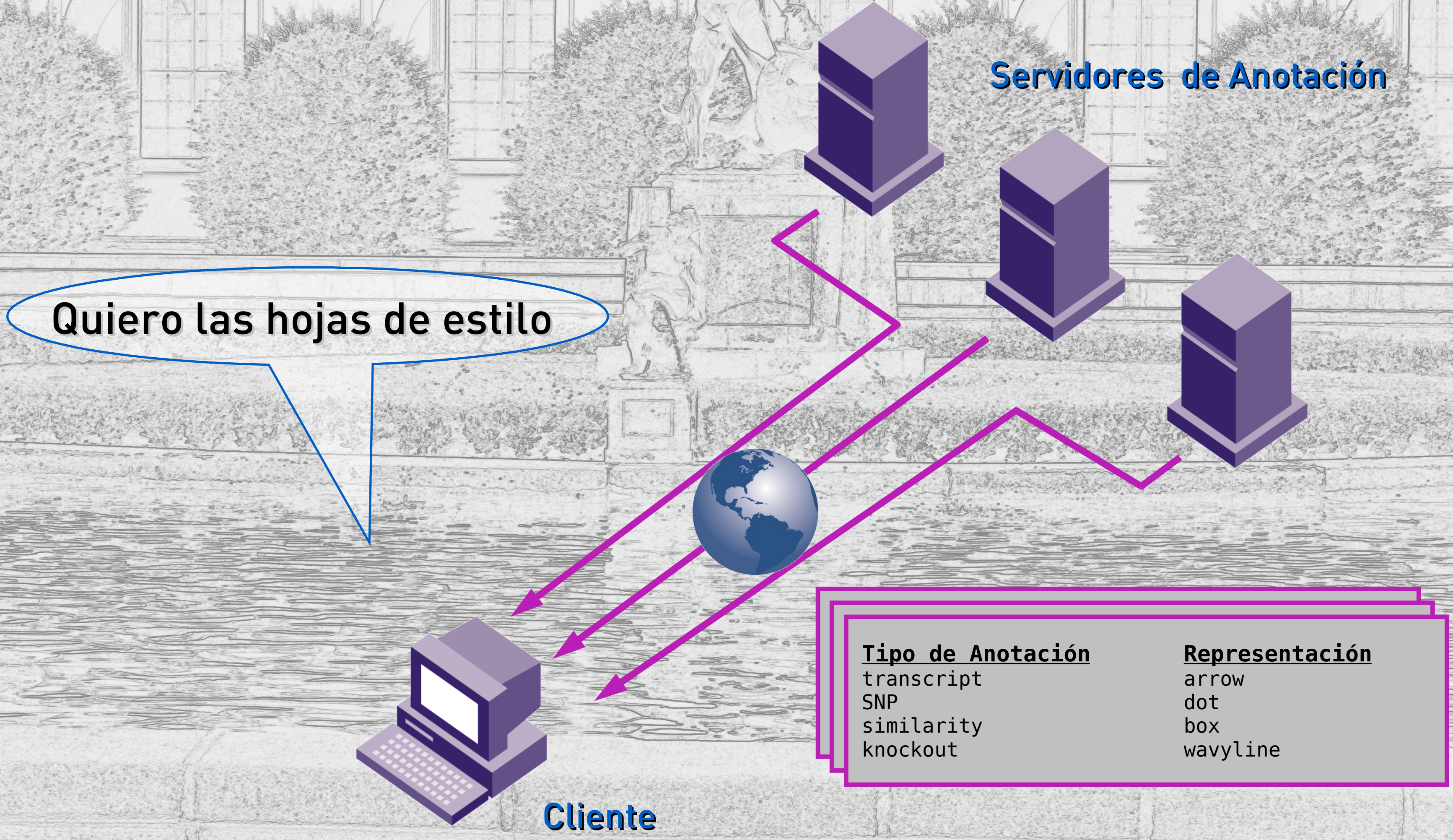
José Manuel Rodríguez Carrasco (jmrodriguez@cnio.es)
José María Fernández González (jmfernandez@cnio.es)

Protocolo DAS
(6) Recuperar anotaciones



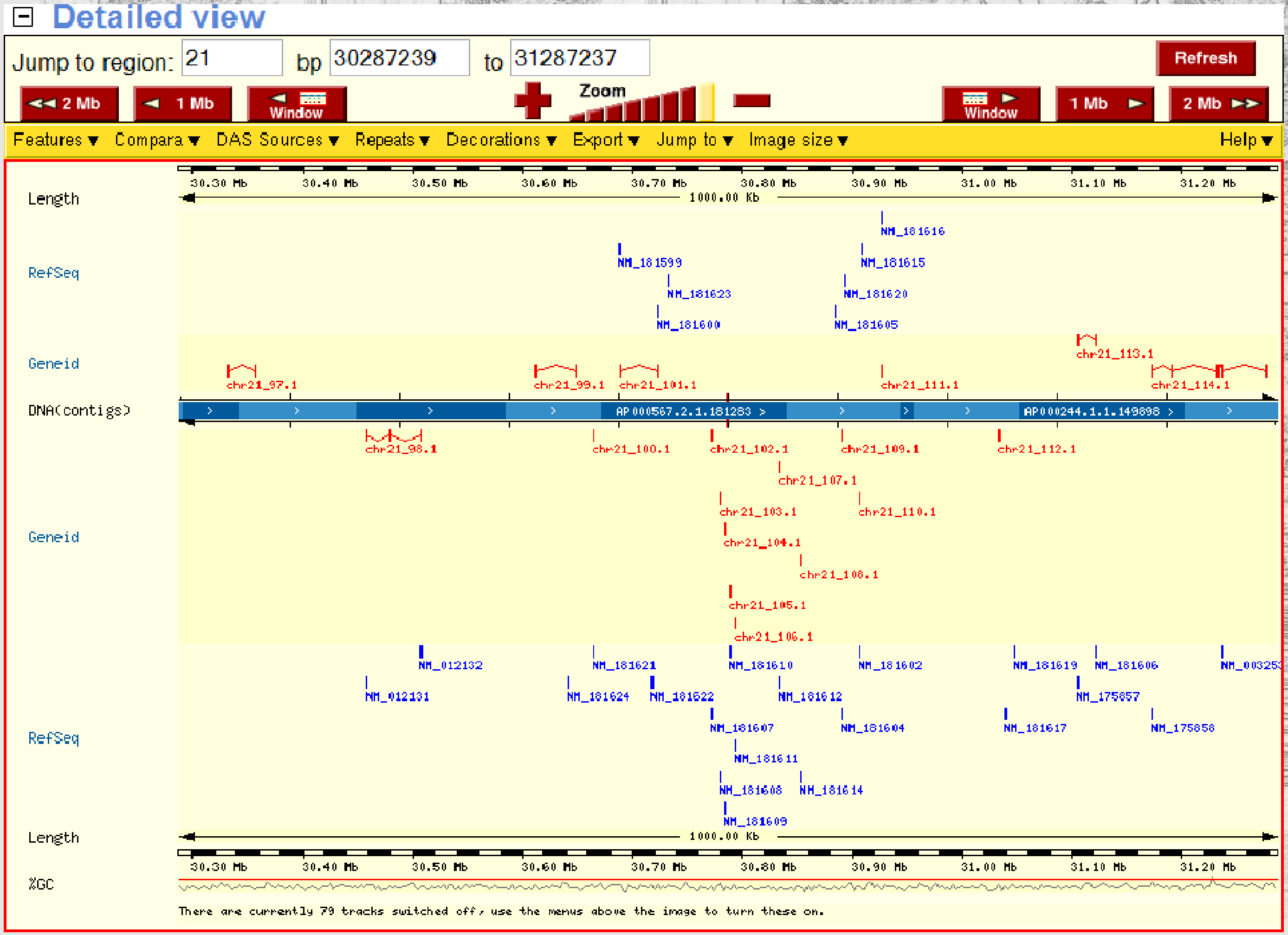
Anotación
ID
Version
Ref sequence
Start, stop
Type, method, category
URL to link to

Protocolo DAS
(7, opcional) Recuperar hojas de estilo



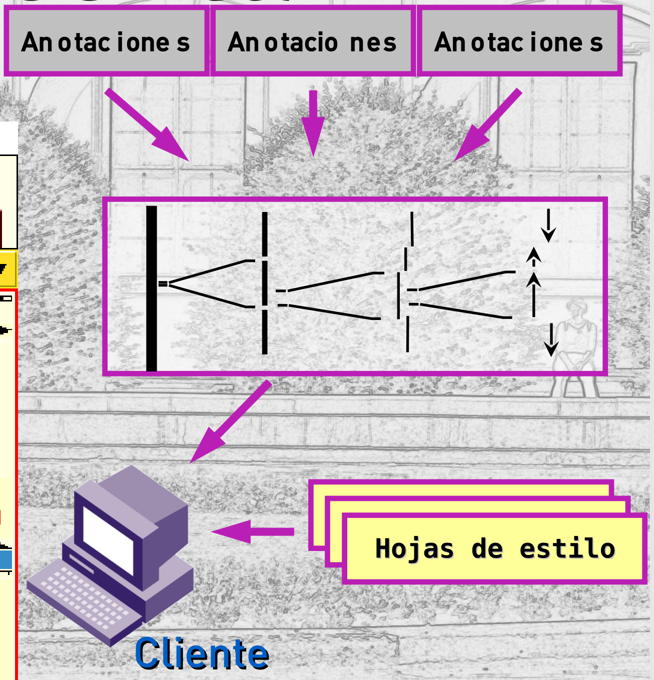
Tipo de Anotación	**Representación**
transcript	arrow
SNP	dot
similarity	box
knockout	wavyline

Protocolo DAS
(8) Integrar y representar



Jump to region: 21 bp 30287239 to 31287237

There are currently 79 tracks switched off, use the menus above the image to turn these on.



Anotaciones

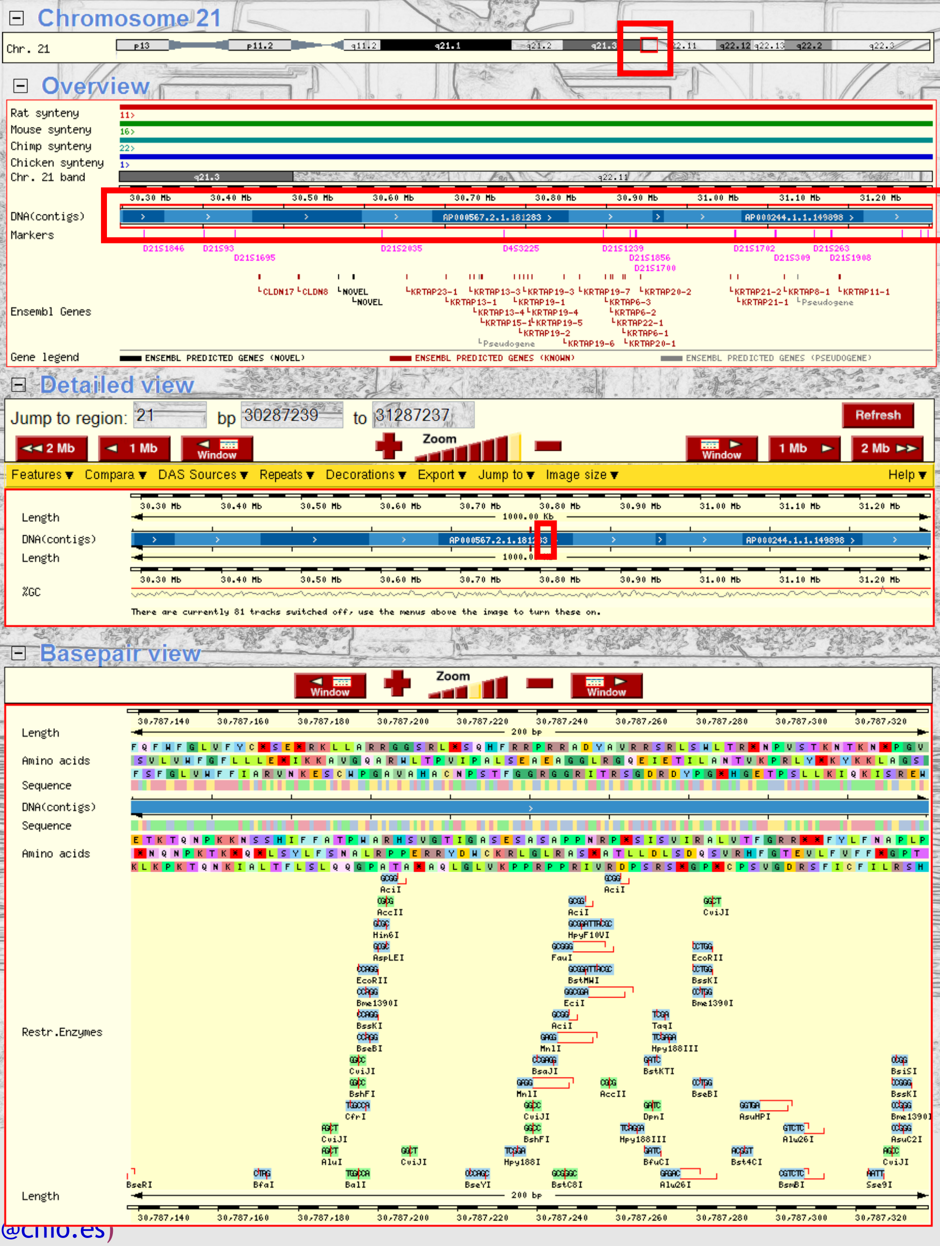
Anotaciones

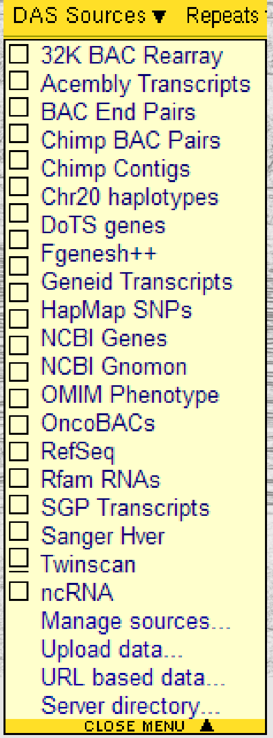
Anotaciones

Hojas de estilo

Cliente

Ejemplo Cliente DAS

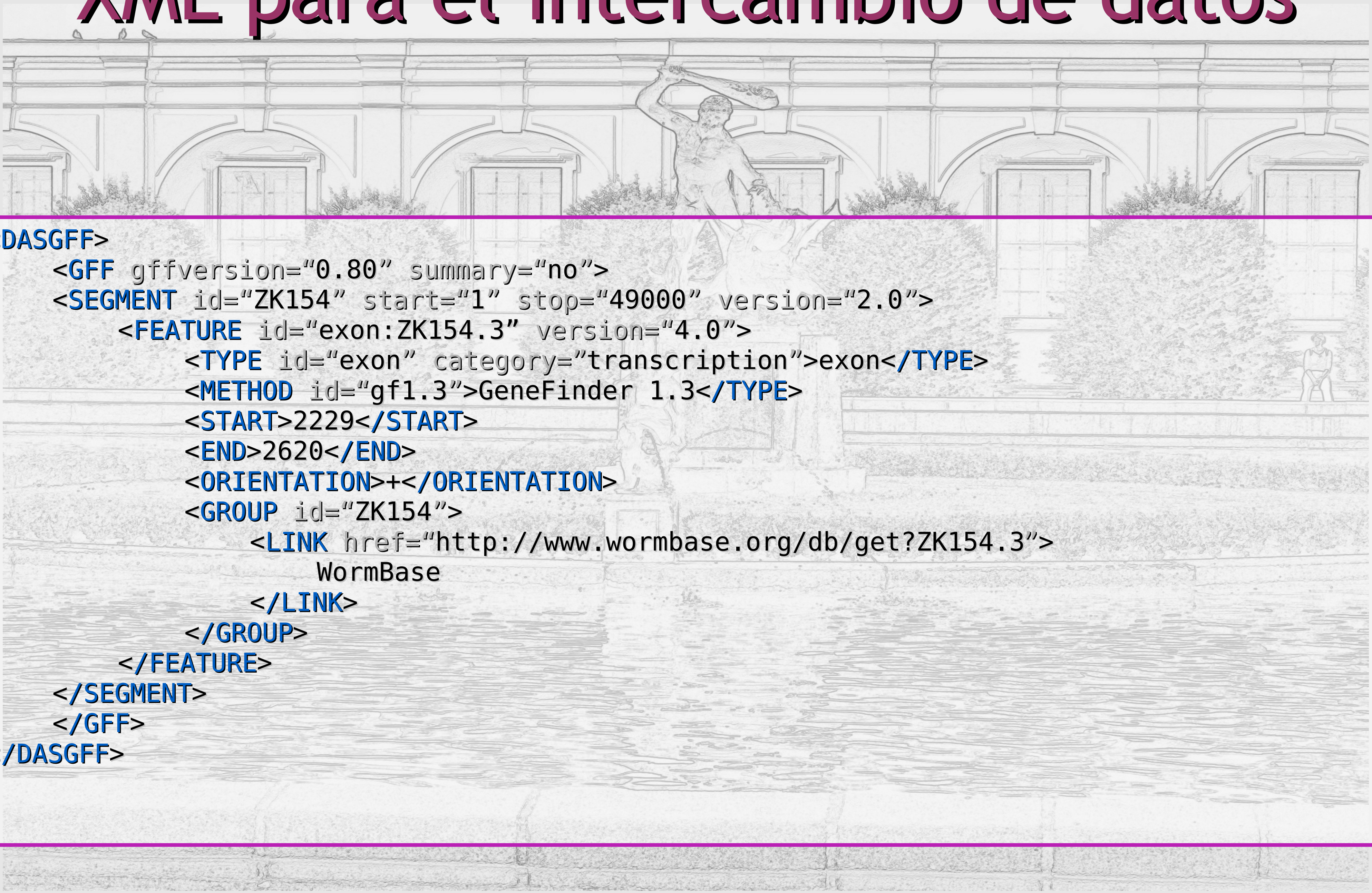




- ☐ 32K BAC Rearray
- ☐ Acembly Transcripts
- ☐ BAC End Pairs
- ☐ Chimp BAC Pairs
- ☐ Chimp Contigs
- ☐ Chr20 haplotypes
- ☐ DoTS genes
- ☐ Fgenesh++
- ☐ Geneid Transcripts
- ☐ HapMap SNPs
- ☐ NCBI Genes
- ☐ NCBI Gnomon
- ☐ OMIM Phenotype
- ☐ OncoBACs
- ☐ RefSeq
- ☐ Rfam RNAs
- ☐ SGP Transcripts
- ☐ Sanger Hver
- ☐ Twinscan
- ☐ ncRNA

Manage sources...
Upload data...
URL based data...
Server directory...

XML para el intercambio de datos

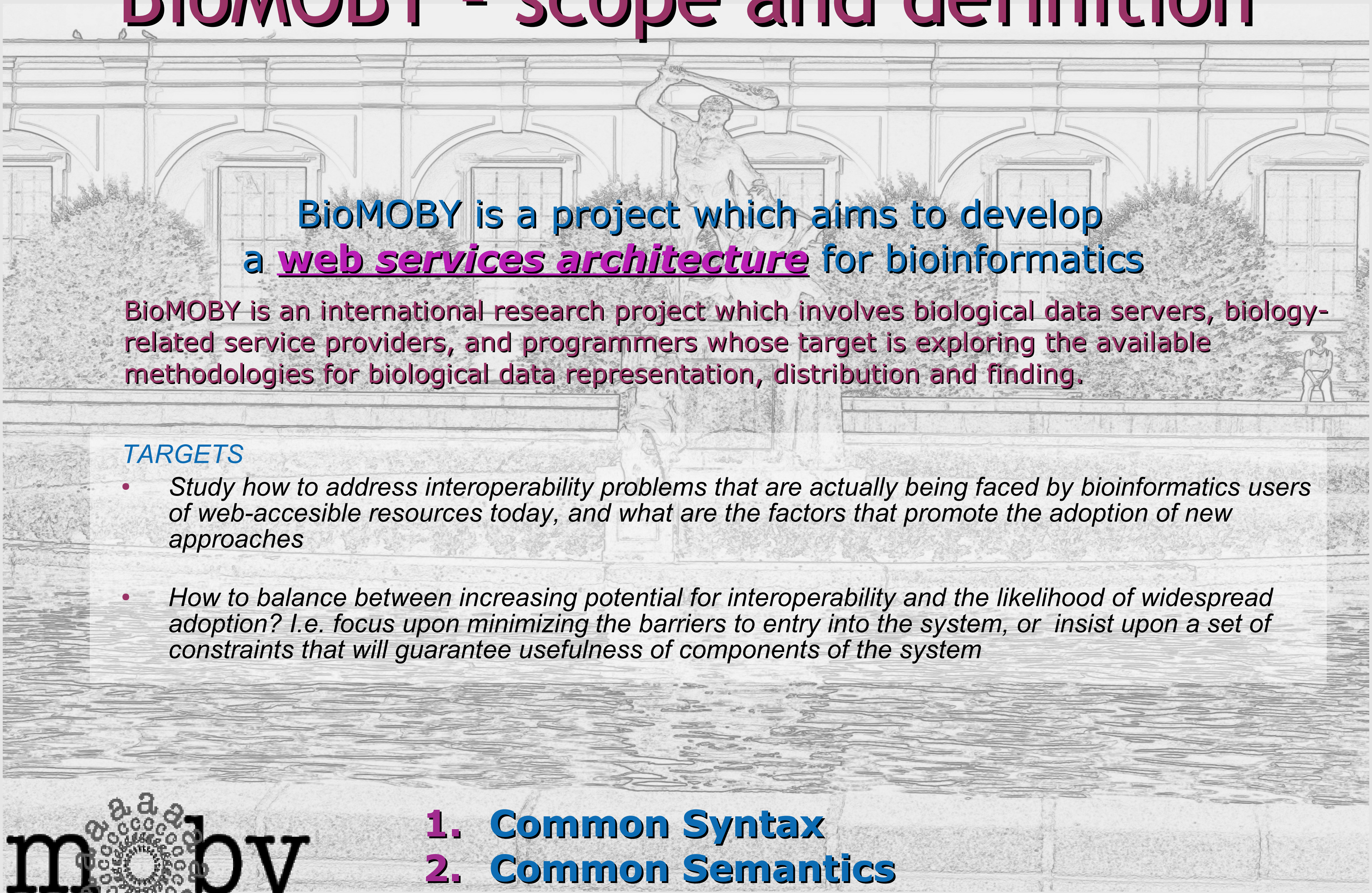


```
<DASGFF>
    <GFF gffversion=“0.80” summary=“no”>
    <SEGMENT id=“ZK154” start=“1” stop=“49000” version=“2.0”>
        <FEATURE id=“exon:ZK154.3” version=“4.0”>
            <TYPE id=“exon” category=”transcription”>exon</TYPE>
            <METHOD id=“gf1.3”>GeneFinder 1.3</TYPE>
            <START>2229</START>
            <END>2620</END>
            <ORIENTATION>+</ORIENTATION>
            <GROUP id=“ZK154”>
                <LINK href=“http://www.wormbase.org/db/get?ZK154.3”>
                    WormBase
                </LINK>
            </GROUP>
        </FEATURE>
    </SEGMENT>
    </GFF>
</DASGFF>
```

WEB SERVICES EN BIOINFORMÁTICA

moby

BioMOBY – scope and definition



BioMOBY is a project which aims to develop a *web services architecture* for bioinformatics

BioMOBY is an international research project which involves biological data servers, biology-related service providers, and programmers whose target is exploring the available methodologies for biological data representation, distribution and finding.

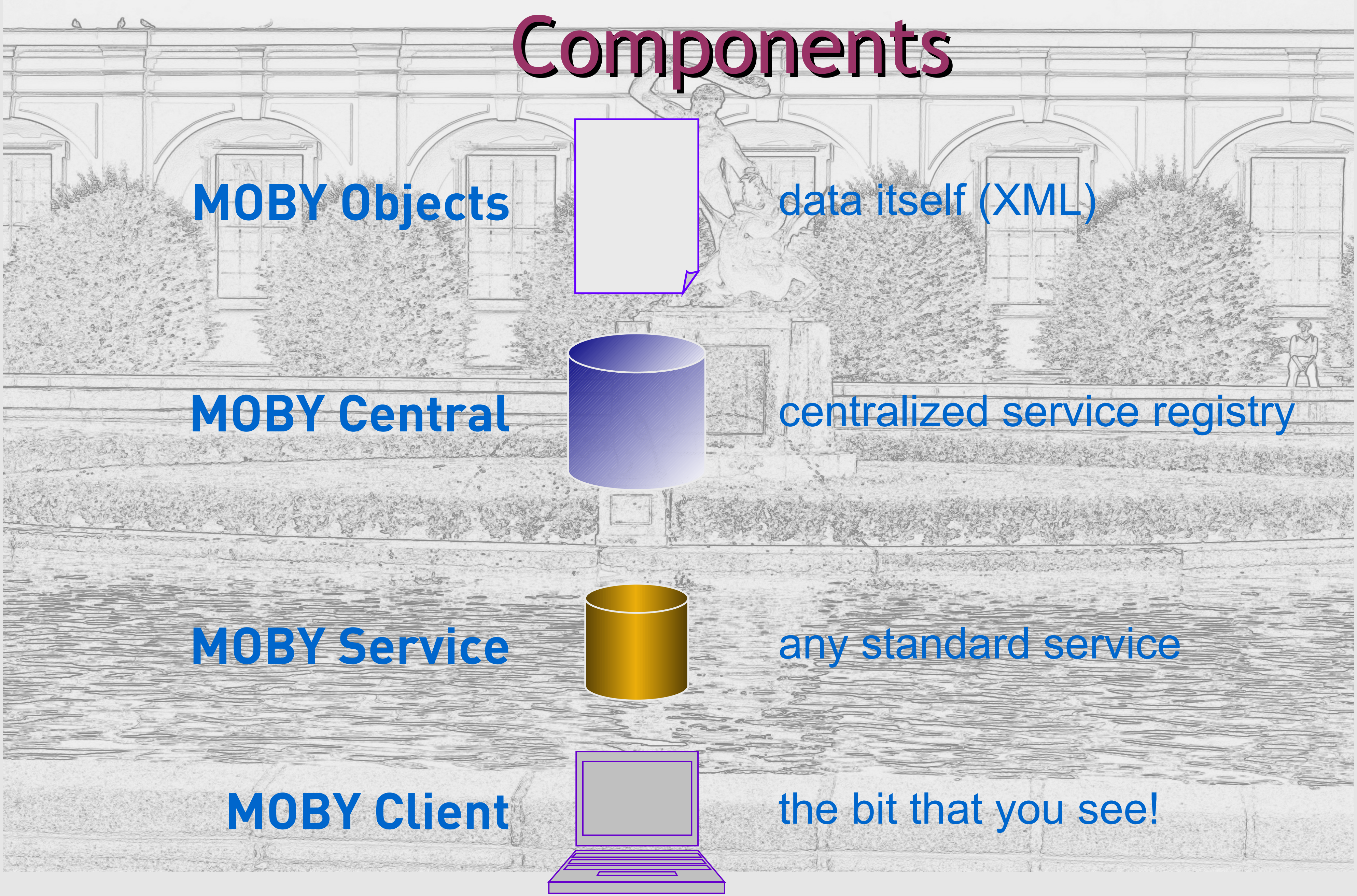
TARGETS

- *Study how to address interoperability problems that are actually being faced by bioinformatics users of web-accesible resources today, and what are the factors that promote the adoption of new approaches*
- *How to balance between increasing potential for interoperability and the likelihood of widespread adoption? I.e. focus upon minimizing the barriers to entry into the system, or insist upon a set of constraints that will guarantee usefulness of components of the system*

1. **Common Syntax**
2. **Common Semantics**
3. ***Dynamic Discovery***

José Manuel Rodríguez Carrasco (jmrodriguez@cnio.es)
José María Fernández González (jmfernandez@cnio.es)

MOBY-S system Components



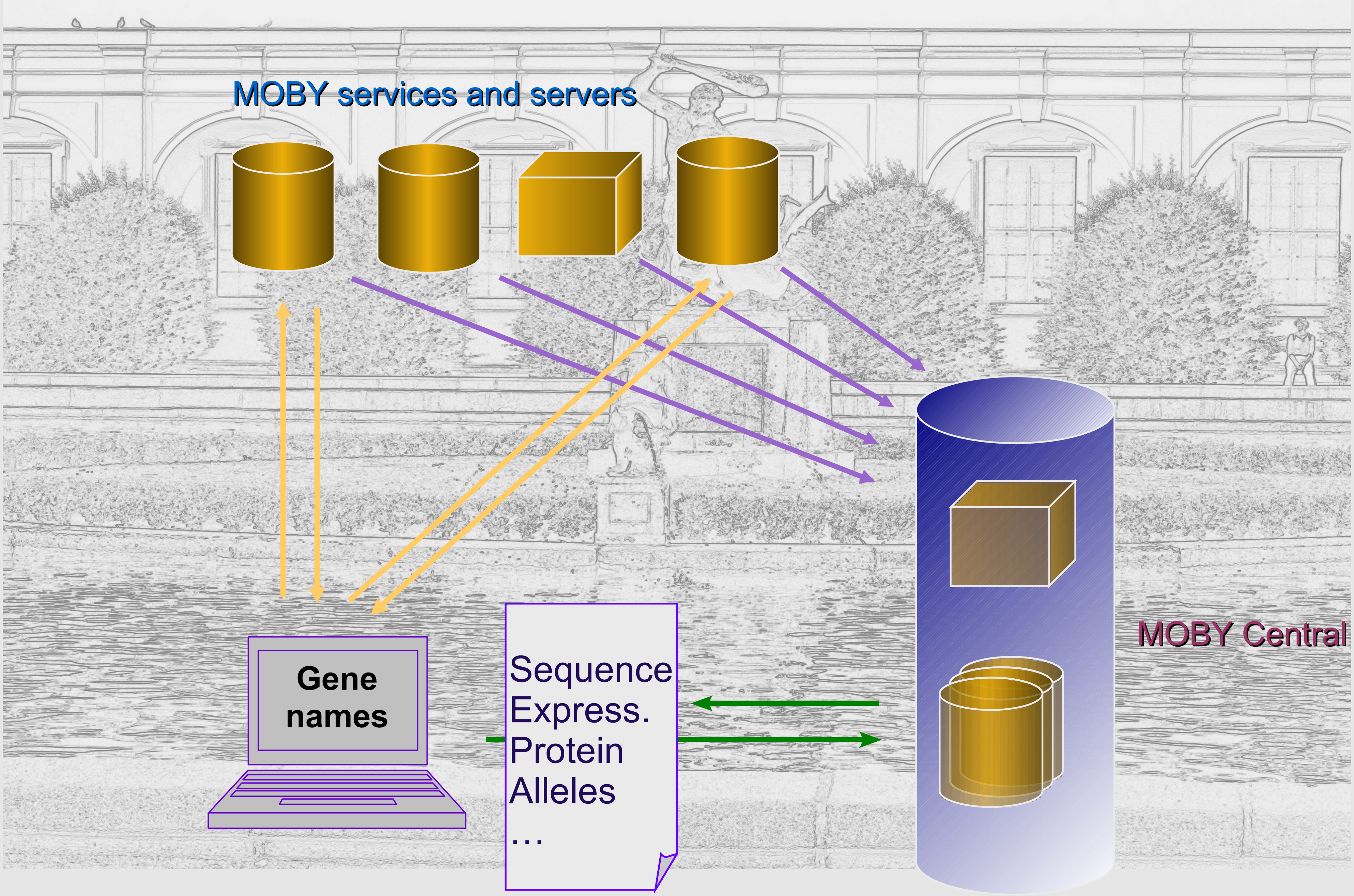
MOBY Objects data itself (XML)

MOBY Central centralized service registry

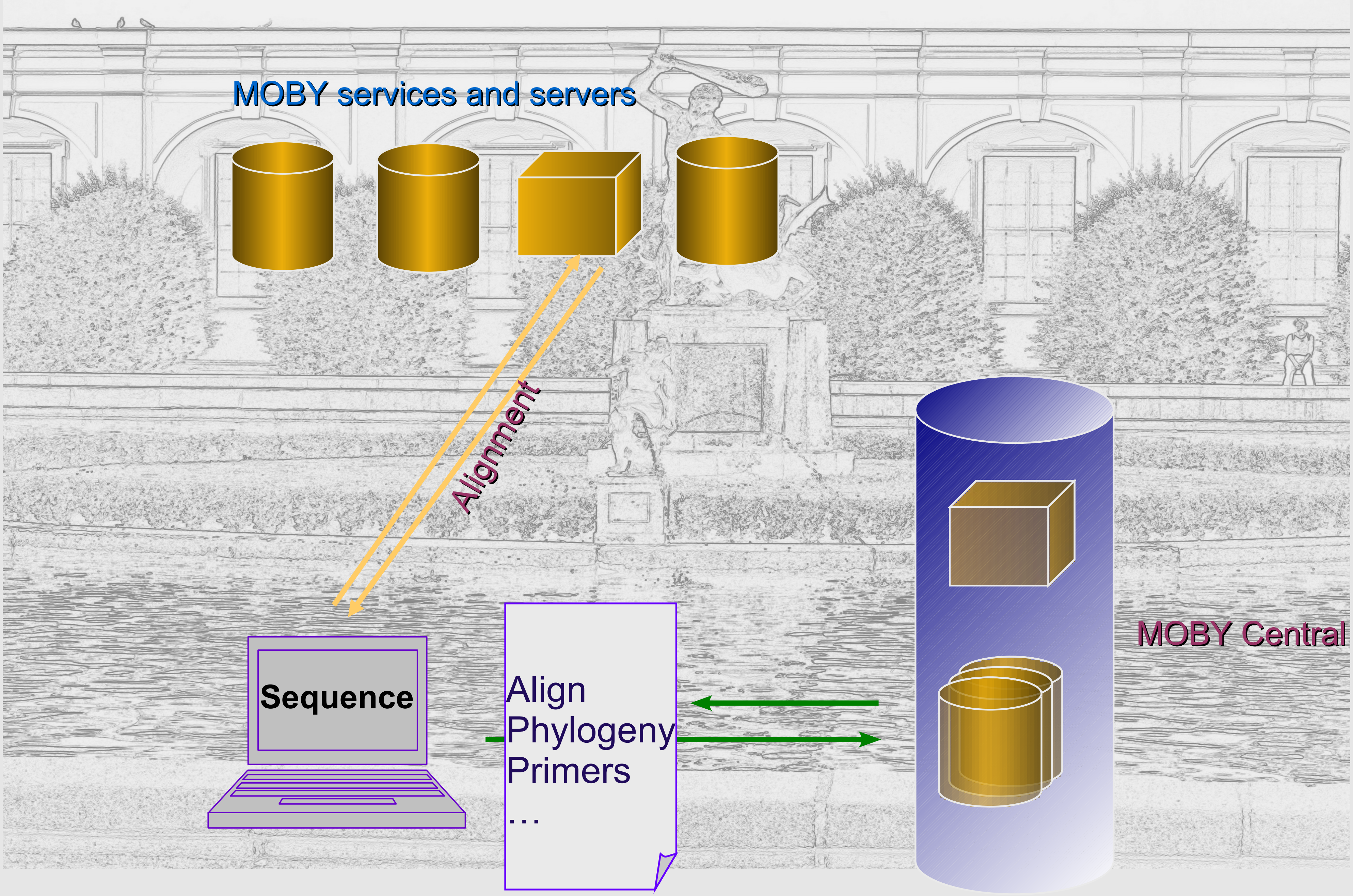
MOBY Service any standard service

MOBY Client the bit that you see!

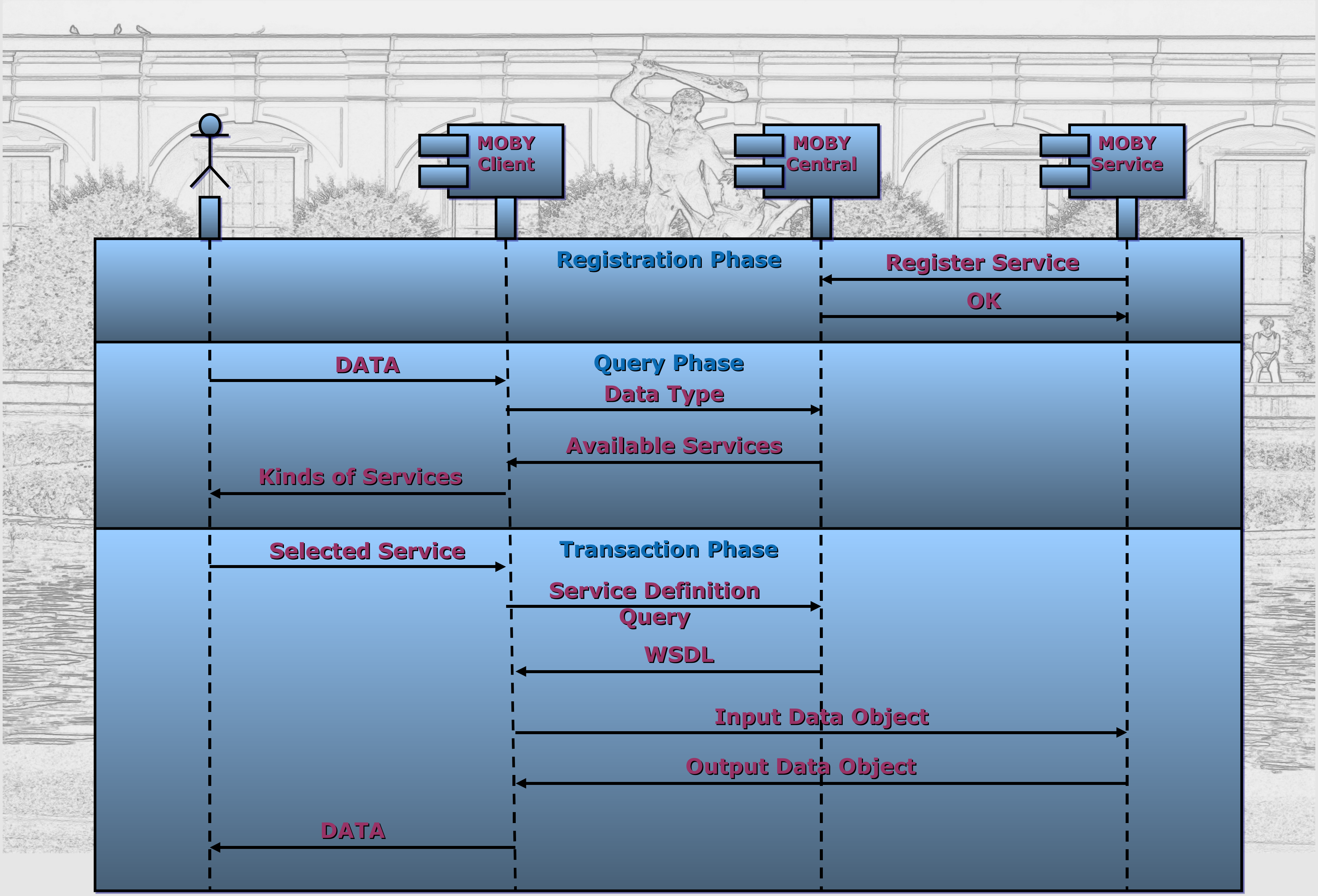
MOBY-S overview (I)



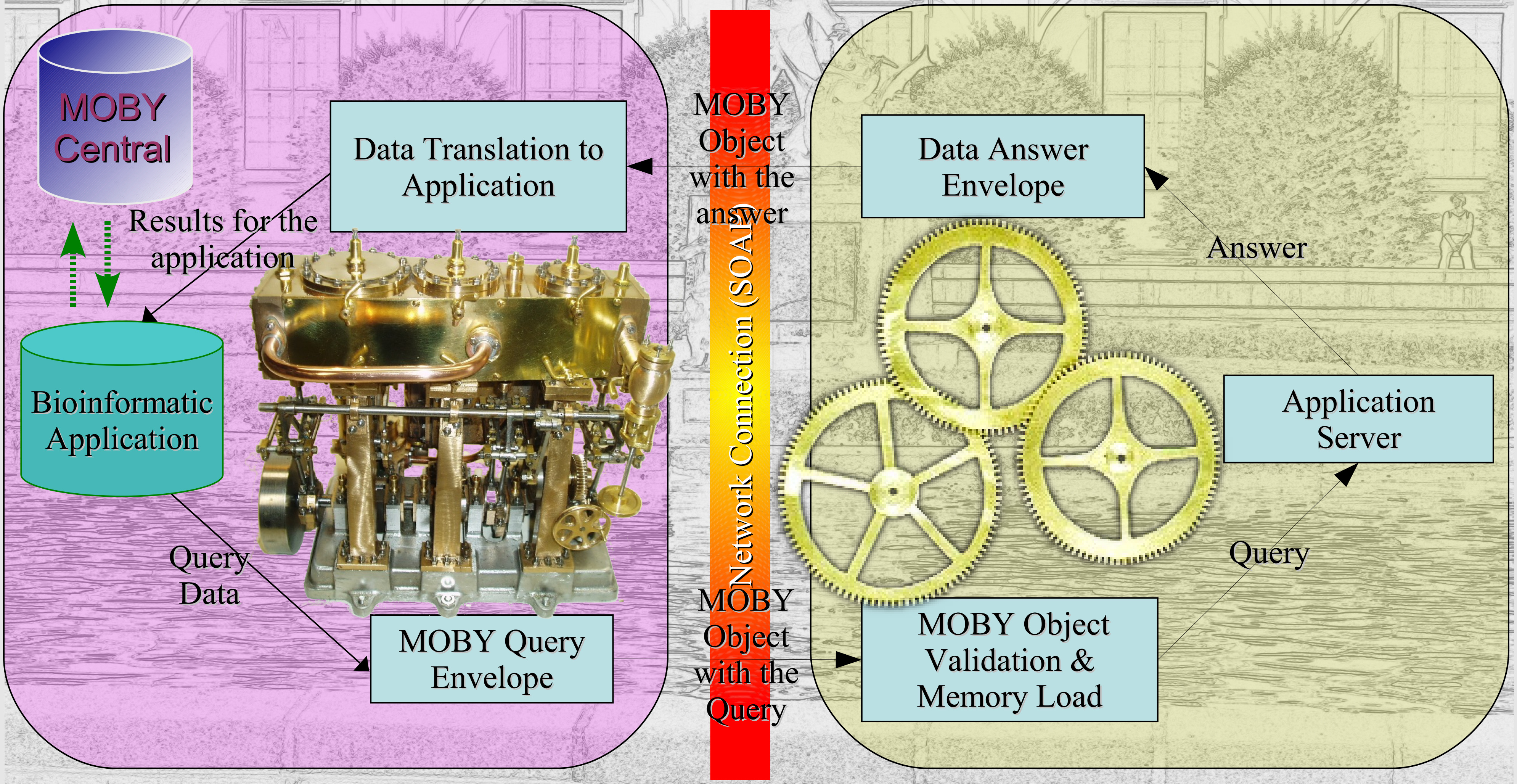
MOBY-S overview (II)



A BioMOBY session



BioMOBY Service Query Processing



MOBY Central

Data Translation to Application

Results for the application

Bioinformatic Application

Query Data

MOBY Query Envelope

MOBY Object with the answer

Network Connection (SOAP)

MOBY Object with the Query

Data Answer Envelope

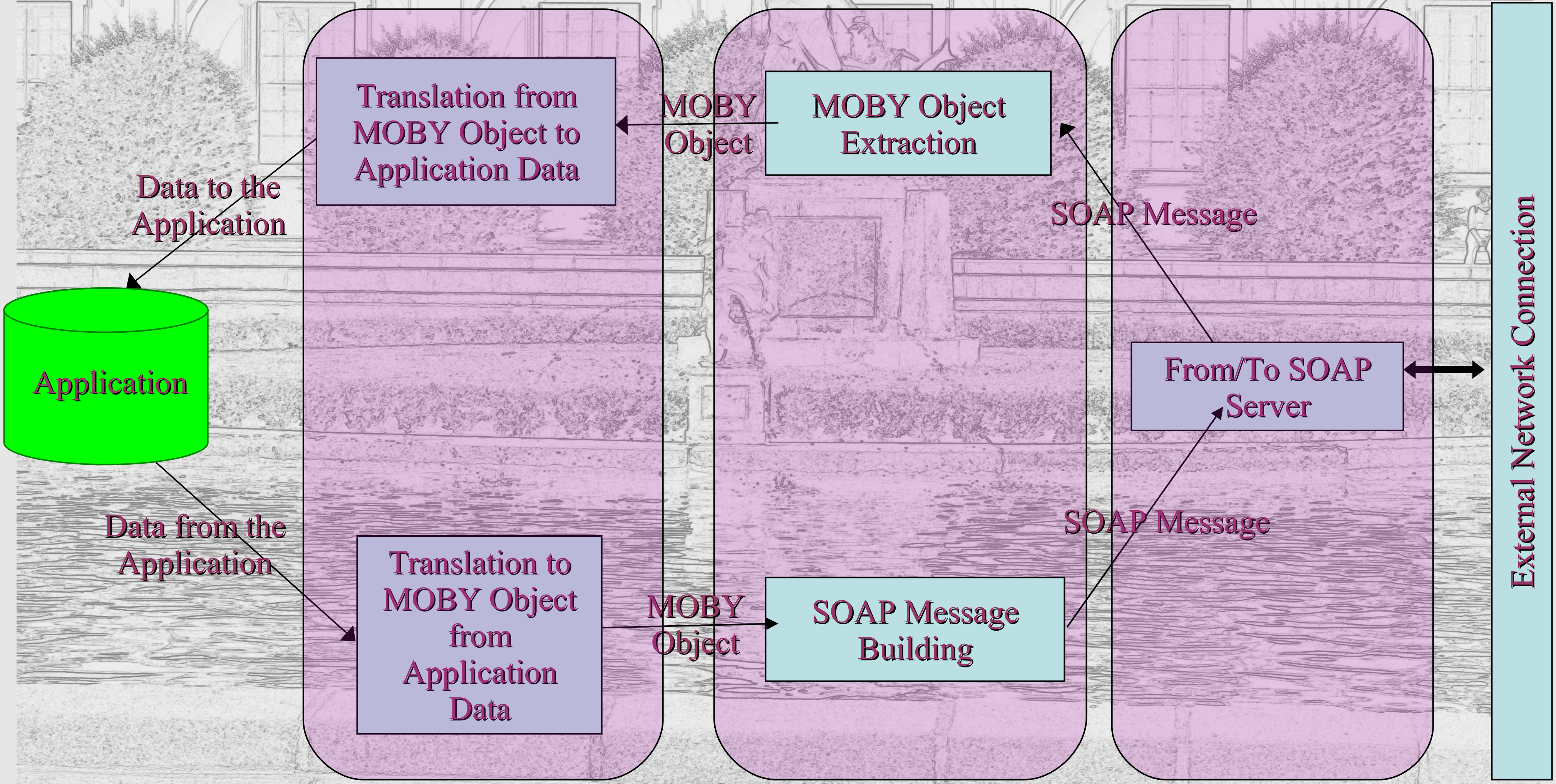
Answer

Application Server

Query

MOBY Object Validation & Memory Load

MOBY messages dispatching

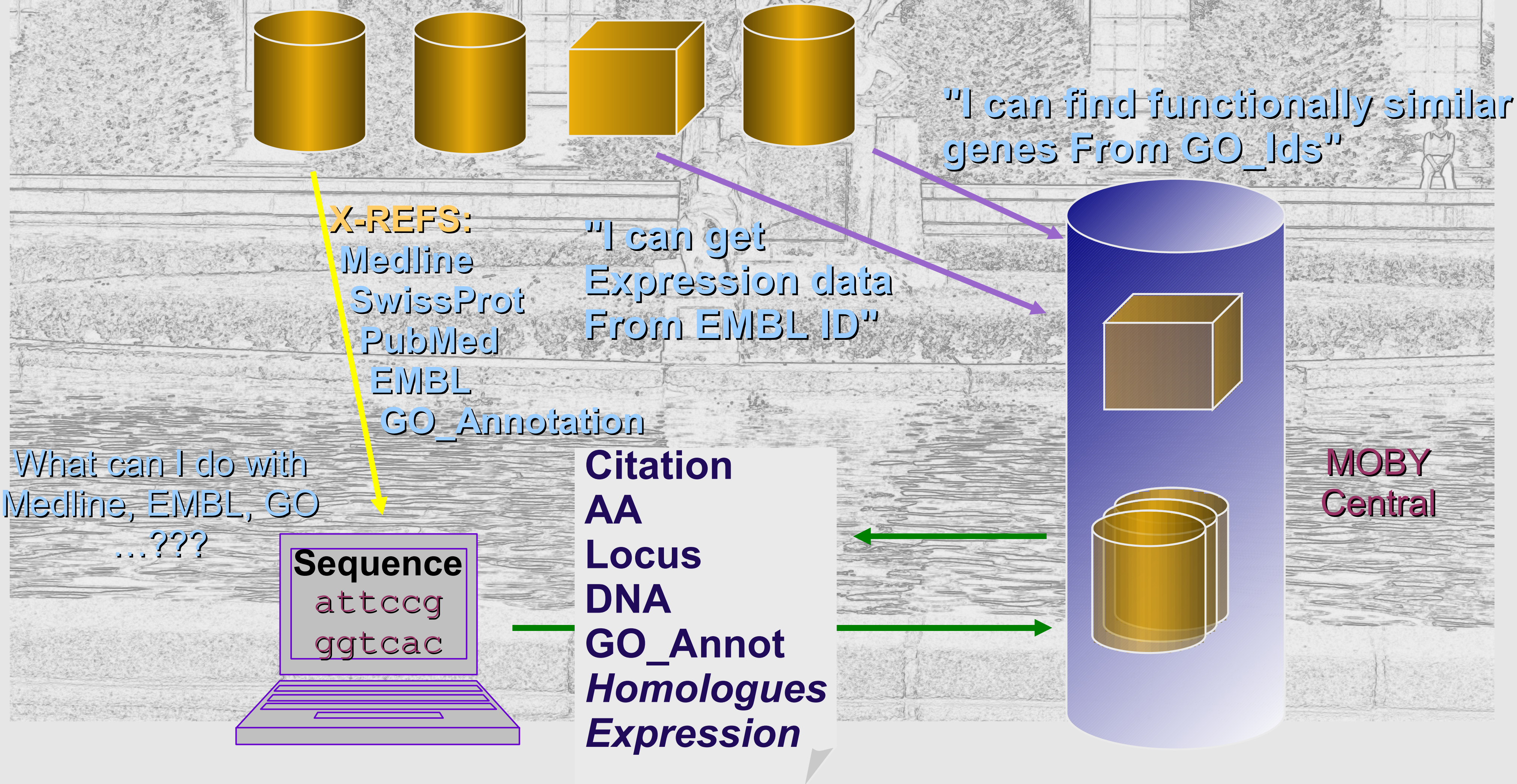


MOBY Service Internal Skeleton

José Manuel Rodríguez Carrasco (jmrodriguez@cnio.es)
José María Fernández González (jmfernandez@cnio.es)

Where is new information from?

“SERVICE AMPLIFICATION”



"I can find functionally similar genes From GO_Ids"

X-REFS:
Medline
SwissProt
PubMed
EMBL
GO_Annotation

"I can get Expression data From EMBL ID"

What can I do with Medline, EMBL, GO …???

Sequence
attccg
ggtcac

Citation
AA
Locus
DNA
GO_Annot
Homologues
Expression

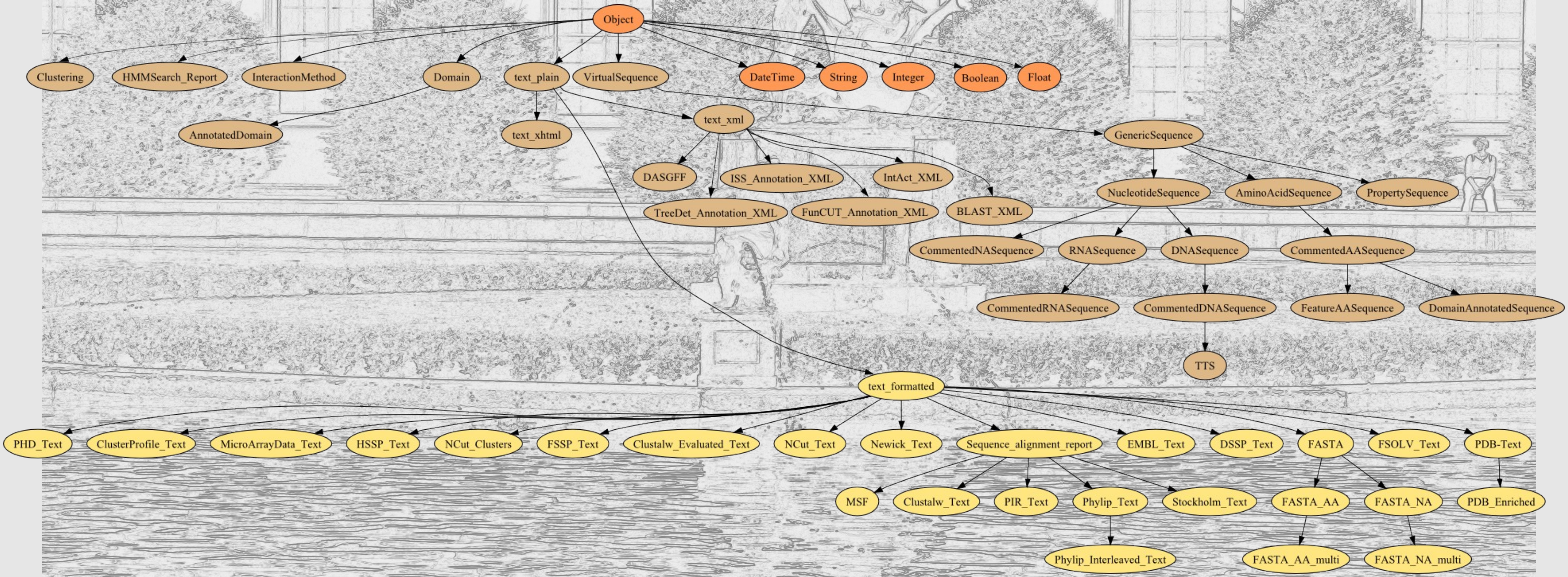
MOBY Central

José Manuel Rodríguez Carrasco (jmrodriguez@cnio.es)
José María Fernández González (jmfernandez@cnio.es)

MOBY-S uses ontologies

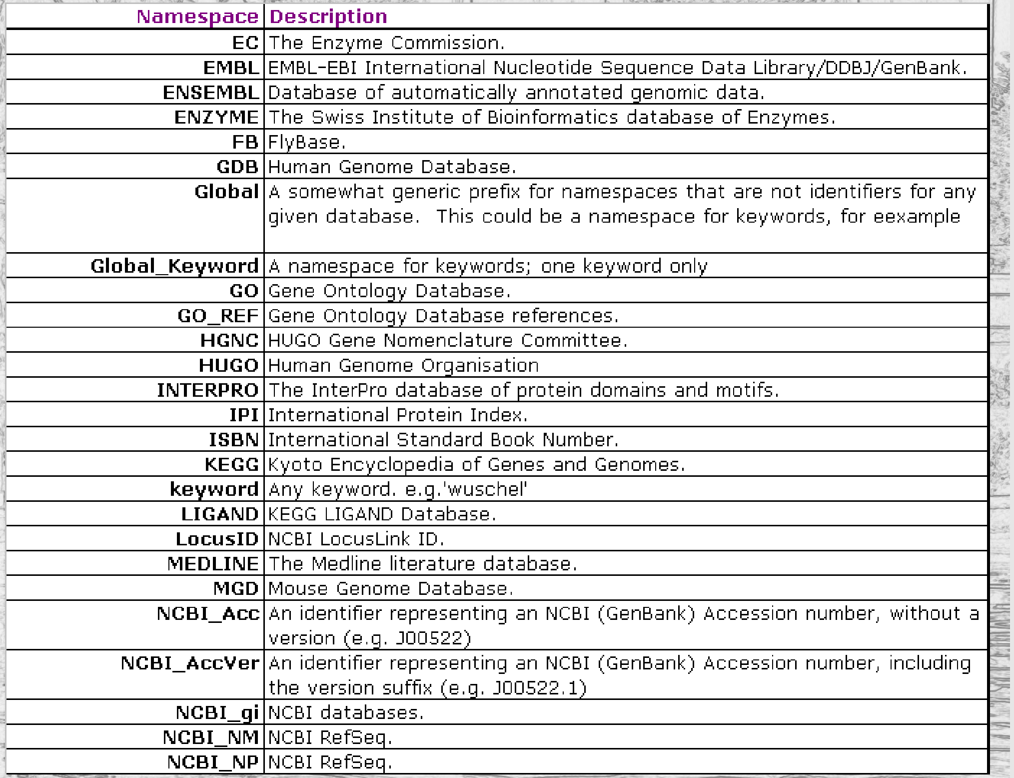
- By classifying data and services we make searching more powerful
- **Object ontology** types of input/output objects
- **Service type ontology** types of services (currently very poor)
  - Analysis - Service analyzes data
    - NCBI-BLAST - Analysis Blast sequences using NCBI Blast program
  - Parsing - Service parses data
  - Registration - Service registers data
  - Retrieval – Service resolves data
  - Resolution - The base retrieval type
- **Service ontology** available services
- **Namespace ontology** scope of the identifiers

A piece of the MOBY Object Ontology



MOBY and Data Namespaces

- Examples: AGI_LocusCode or NCBI_gi
- MOBY adopted *Gene Ontology Cross-reference Abbreviations List* as its namespace list.
- Namespace list is not closed to the previous ones: you can add the ones you need.
- Each input parameter of the services which are registered at MOBY Central is usually restricted to one or more namespaces.
- E.g. If a service takes as input parameter a GenericSequence object from any namespace, then it is registered with no restriction about the namespaces that input parameter can take.



Namespace	Description
EC	The Enzyme Commission.
EMBL	EMBL-EBI International Nucleotide Sequence Data Library/DDBJ/GenBank.
ENSEMBL	Database of automatically annotated genomic data.
ENZYME	The Swiss Institute of Bioinformatics database of Enzymes.
FB	FlyBase.
GDB	Human Genome Database.
Global	A somewhat generic prefix for namespaces that are not identifiers for any given database. This could be a namespace for keywords, for eexample
Global_Keyword	A namespace for keywords; one keyword only
GO	Gene Ontology Database.
GO_REF	Gene Ontology Database references.
HGNC	HUGO Gene Nomenclature Committee.
HUGO	Human Genome Organisation
INTERPRO	The InterPro database of protein domains and motifs.
IPI	International Protein Index.
ISBN	International Standard Book Number.
KEGG	Kyoto Encyclopedia of Genes and Genomes.
keyword	Any keyword. e.g.'wuschel'
LIGAND	KEGG LIGAND Database.
LocusID	NCBI LocusLink ID.
MEDLINE	The Medline literature database.
MGD	Mouse Genome Database.
NCBI_Acc	An identifier representing an NCBI (GenBank) Accession number, without a version (e.g. J00522)
NCBI_AccVer	An identifier representing an NCBI (GenBank) Accession number, including the version suffix (e.g. J00522.1)
NCBI_gi	NCBI databases.
NCBI_NM	NCBI RefSeq.
NCBI_NP	NCBI RefSeq.

MOBY Objects

Each MOBY Object is made by a triplet:
object name, identificator and namespace.
But it also has:

Payload (it brings additional data about the object)

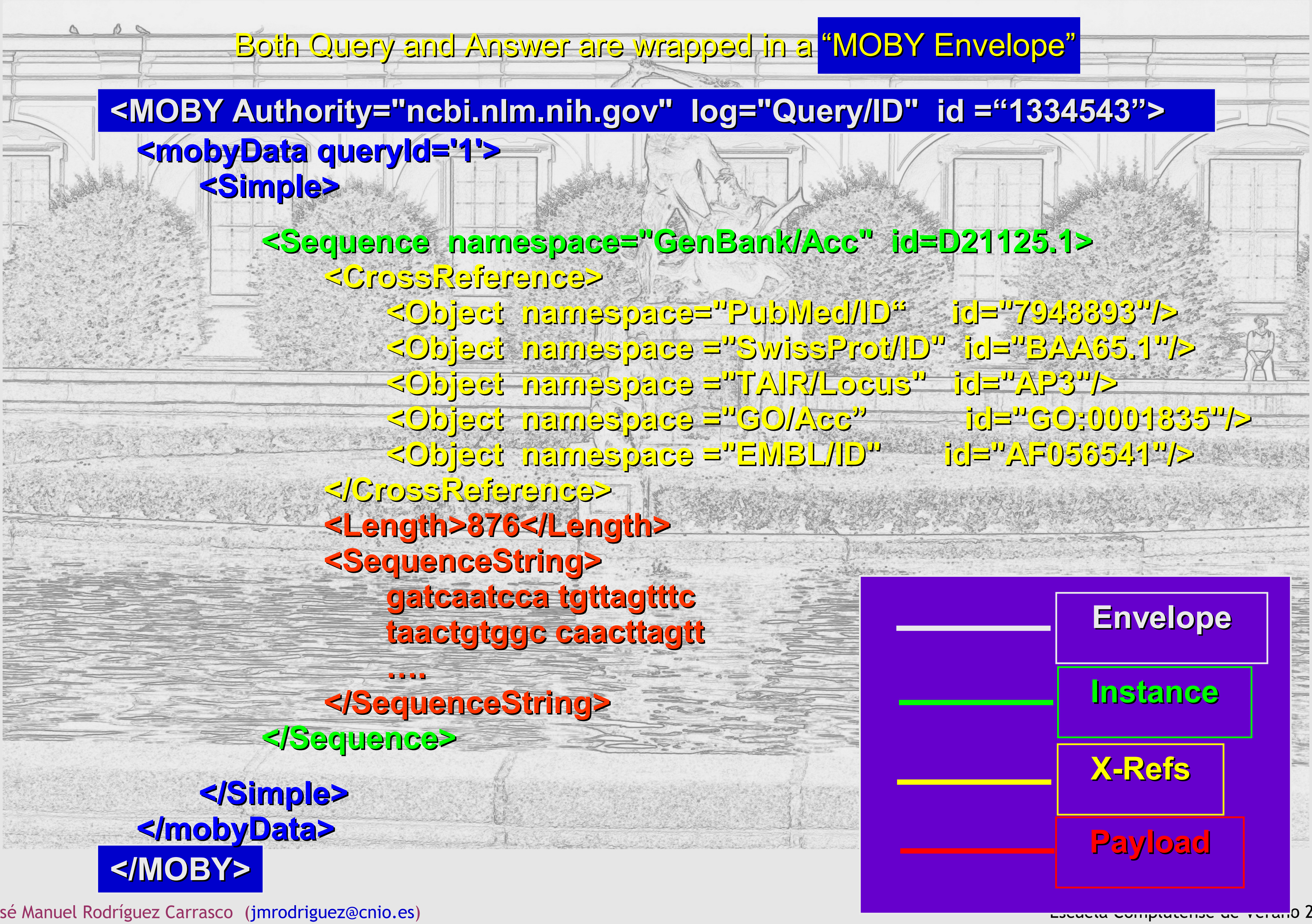
Cross-Reference Information Block (CRIB)

```
<Sequence  namespace="GenBank/Acc"  id=D21125.1>
    <CrossReference>
        <Object  namespace="PubMed/ID“     id="7948893"/>
        <Object  namespace ="SwissProt/ID"  id="BAA65.1"/>
        <Object  namespace ="TAIR/Locus"   id="AP3"/>
        <Object  namespace ="GO/Acc”          id="GO:0001835"/>
        <Object  namespace ="EMBL/ID"      id="AF056541"/>

   </CrossReference>
   <Length>876</Length>
       <SequenceString>
               gatcaatcca tgttagtttc
               taactgtggc caacttagtt
                ….
       </SequenceString>
</Sequence>
```

“Light” object inheritance
based on simpler objects
It depends on the instance

MOBY Queries



Both Query and Answer are wrapped in a "MOBY Envelope"

```
<MOBY Authority="ncbi.nlm.nih.gov"  log="Query/ID"  id ="1334543">
   <mobyData queryId='1'>
        <Simple>

             <Sequence  namespace="GenBank/Acc"  id=D21125.1>
                  <CrossReference>
                       <Object  namespace="PubMed/ID"     id="7948893"/>
                       <Object  namespace ="SwissProt/ID"  id="BAA65.1"/>
                       <Object  namespace ="TAIR/Locus"   id="AP3"/>
                       <Object  namespace ="GO/Acc"          id="GO:0001835"/>
                       <Object  namespace ="EMBL/ID"      id="AF056541"/>
                  </CrossReference>
                  <Length>876</Length>
                  <SequenceString>
                       gatcaatcca tgttagtttc
                       taactgtggc caacttagtt
                       ....
                  </SequenceString>
             </Sequence>

        </Simple>
   </mobyData>
</MOBY>
```

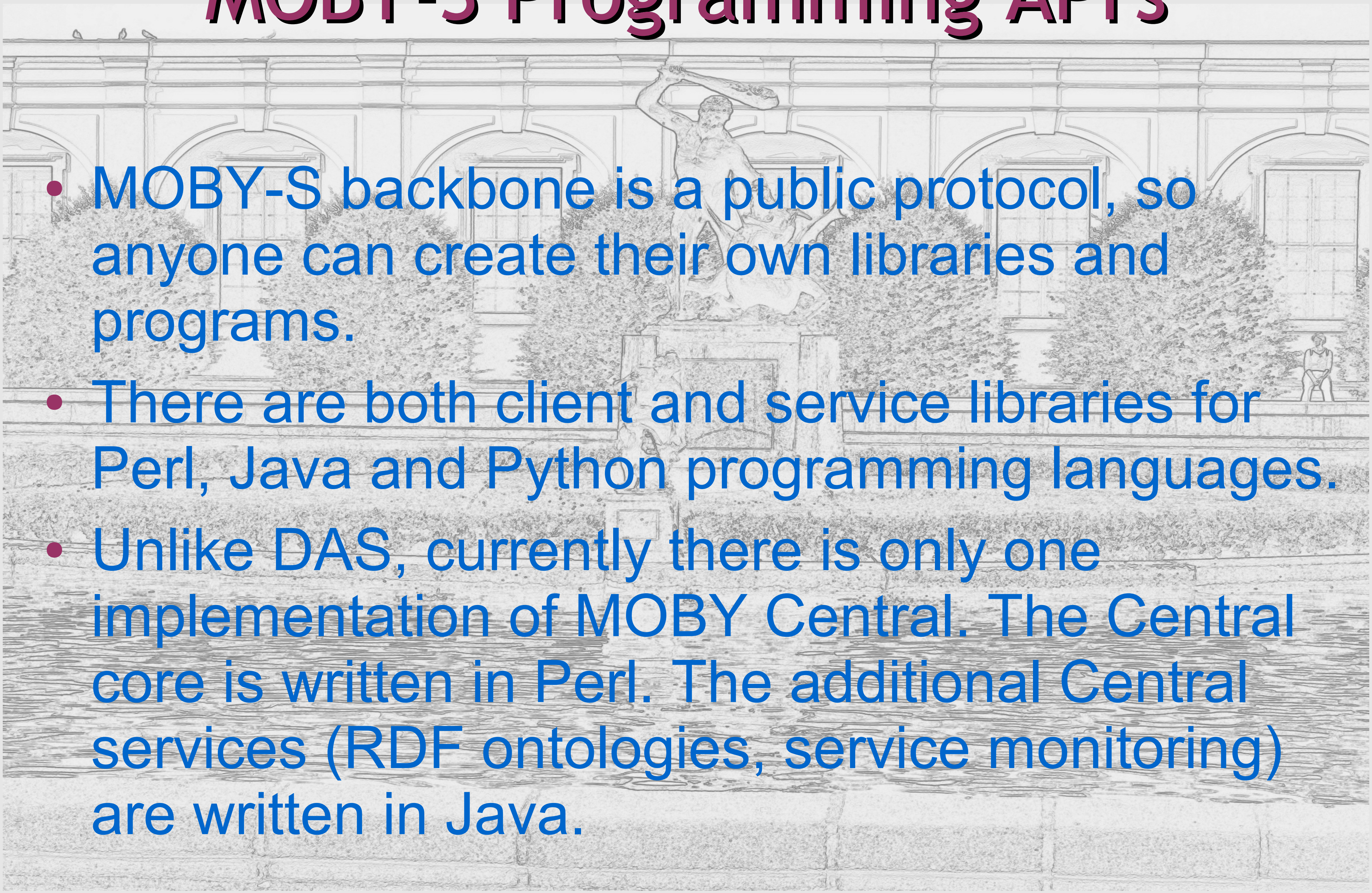
Envelope

Instance

X-Refs

Payload

MOBY-S Programming API's



- MOBY-S backbone is a public protocol, so anyone can create their own libraries and programs.
- There are both client and service libraries for Perl, Java and Python programming languages.
- Unlike DAS, currently there is only one implementation of MOBY Central. The Central core is written in Perl. The additional Central services (RDF ontologies, service monitoring) are written in Java.

José Manuel Rodríguez Carrasco (jmrodriguez@cnio.es)
José María Fernández González (jmfernandez@cnio.es)

MOBY Service Clients

These clients will allow you to discover and access data and analysis services provided through the BioMOBY framework. Some also allow the construction and execution of re-usable workflows and analytical pipelines.

Client Title	Source	Platform & Language	Description
Gbrowse_ moby	Mark Wilkinson	Web-based, cross-platform	The first Moby client ever written; very simplistic and limited in power.
Taverna	Tom Oinn, Martin Senger; myGrid	Java, cross-platform	Taverna allows linking of inputs and outputs from both Moby and non-Moby services into extensive and complex workflows.
MOWserv	Instituto Nacional de Bioinformatica	Web based, cross-platform	MOWserv allows construction and execution of workflows through a web interface
Remora	Genopole Toulouse	Web based, cross-platform	Remora allows design of Moby workflows through a web interface
Ahab	Benjamin Good, Clarence Kwan, Wilkinson Laboratory, UBC	Web based, cross-platform	Ahab allows parallel execution of multiple services simulteneously.
Seahawk	Paul Gordon, Sun Centre of Excellence, University of Calgary, Alberta, Canada	Java applet, cross-platform	Allows a user to load text or HTML data sources, then discover and execute MOBY Services through hyperlinks and text highlighting.

Clients with embedded MOBY functionality

These are applications that provide a variety of bioinformatics functionality and take advantage of BioMOBY to extend that functionality.

BioTrawler	Frank Gibbons, Harvard	Web based, cross-platform	Explore protein interaction networks by surfing MOBY!
BlueJay	Paul Gordon, University of Calgary (Genome Canada/Genome Alberta)	Web based, cross-platform	Explore genomes with MOBY out-linking functionality
BioFloWeb	Sophie Durand, INRA, France & PlaNet Consortium	Web based, cross-platform	Arabidopsis ‘gene cards’ via PlaNet MOBY services
AtiDB Client	Sean Walsh, John Innes Centre, UK	Web based, cross-platform	Arabidopsis Locus Report via PlaNet MOBY services

These clients allow exploration of the Moby Central registry, e.g. by graph traversal.

Moby Graphs - Martin Senger @ EBI/myGrid
Moby Service Encyclopedia - Ken Stuebe
Registry Browser - Mark Wilkinson

MOBY Constructor Clients

These clients enable exporation, maniputation, and/or editing of the various Moby Ontologies, and entering new service instances into the MOBY Central registry.

Object Ontology

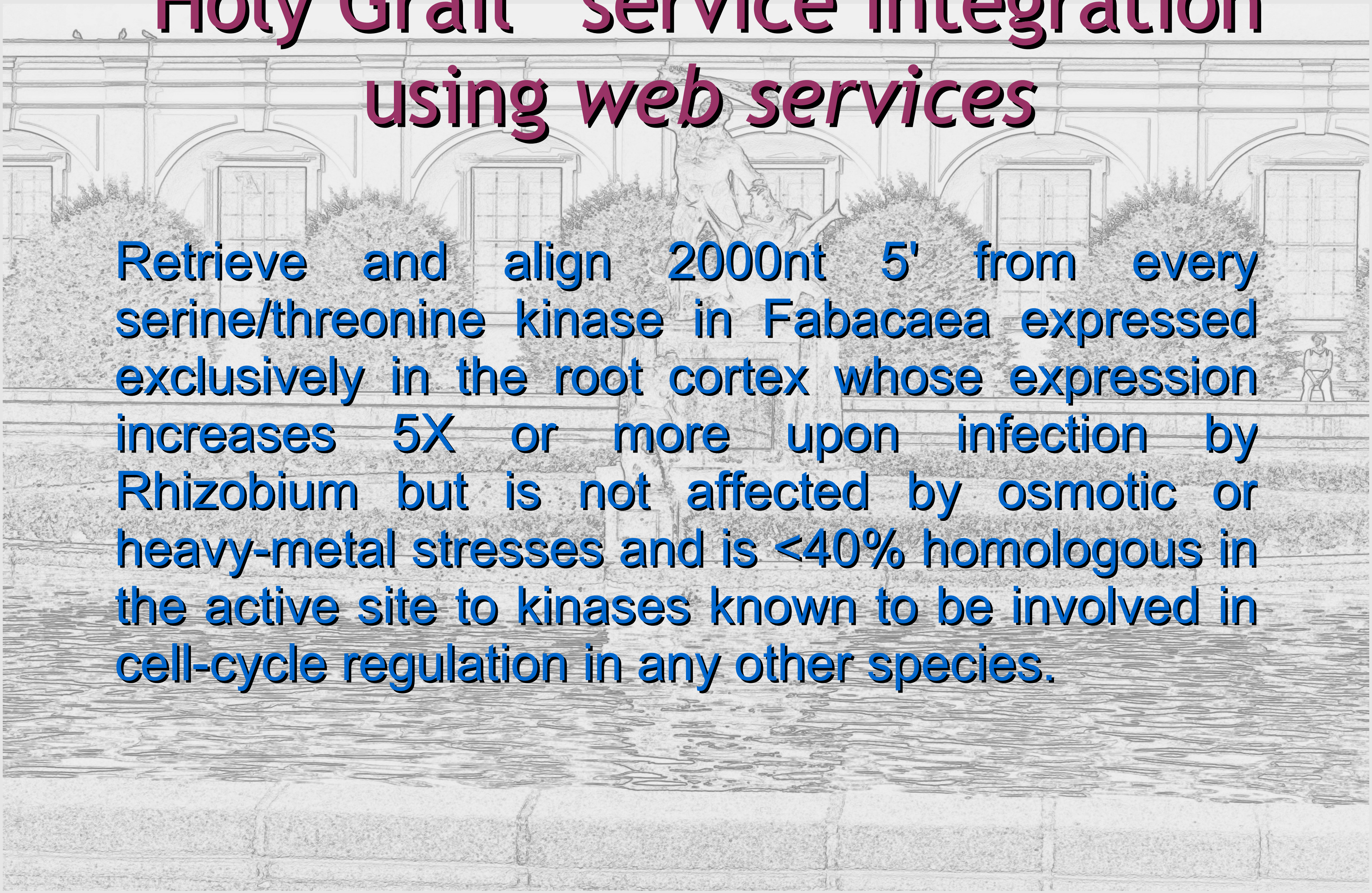
Object Creation Applet - Eddie Kawas, UBC
Object Ontology Browser for Unix and Windows - Yan Wong (un: Invite; pw: Invite)
Object Ontology Browser for OS-X - Yan Wong (un: Invite; pw: Invite)

Service Ontology
Service Type Creation Applet - Eddie Kawas, UBC

Namespace Ontology
Namespace Creation Tool - Ambrose Ng, UBC

MOBY Central - New Service Instance
Service Type Creation Applet - Eddie Kawas, UBC

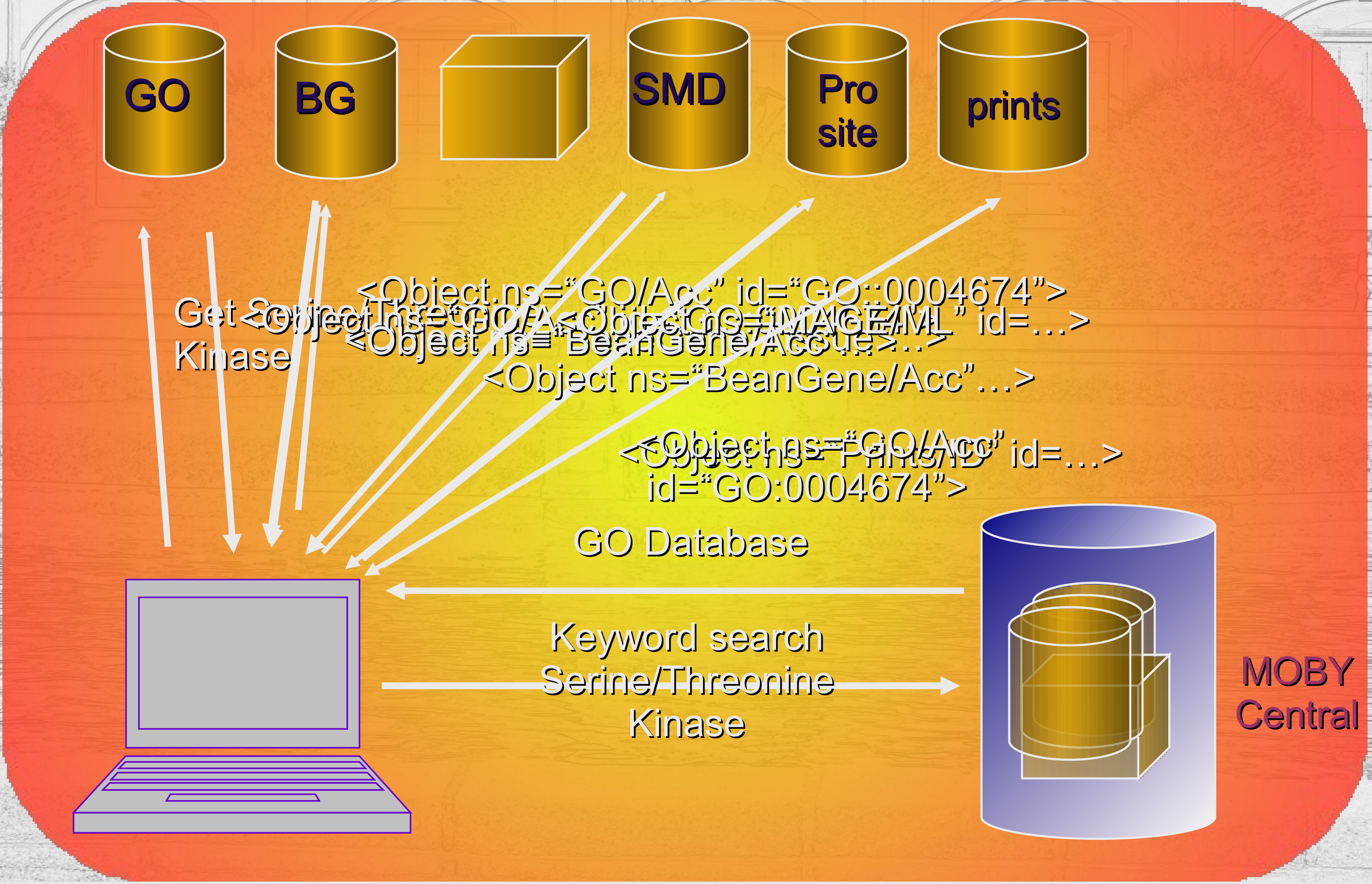
“Holy Grail” service integration using *web services*



Retrieve and align 2000nt 5' from every serine/threonine kinase in Fabacaea expressed exclusively in the root cortex whose expression increases 5X or more upon infection by Rhizobium but is not affected by osmotic or heavy-metal stresses and is <40% homologous in the active site to kinases known to be involved in cell-cycle regulation in any other species.

José Manuel Rodríguez Carrasco (jmrodriguez@cnio.es)
José María Fernández González (jmfernandez@cnio.es)

Workflow which solves the query!



Get Serine/Threonine Kinase

<Object.ns="GO/Acc" id="GO::0004674">

<Object ns="GO/Acc" …

<Object ns="MAGE-ML" id=…>

<Object ns="BeanGene/Accvue>…>

<Object ns="BeanGene/Acc"…>

<Object ns="GO/Acc" id="GO:0004674">

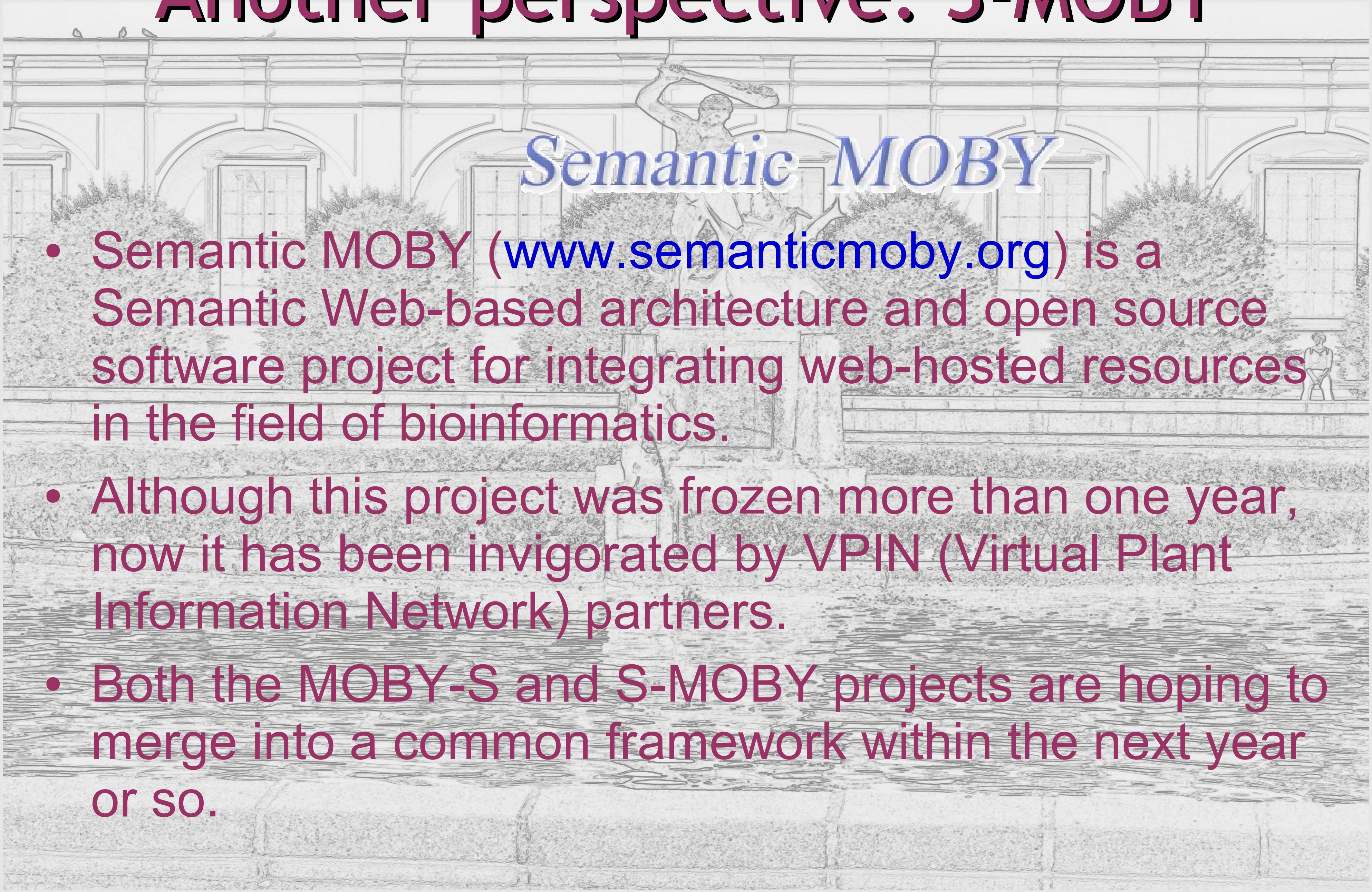
<Object ns="Prints/ID" id=…>

GO Database

Keyword search Serine/Threonine Kinase

José Manuel Rodríguez Carrasco (jmrodriguez@cnio.es)
José María Fernández González (jmfernandez@cnio.es)

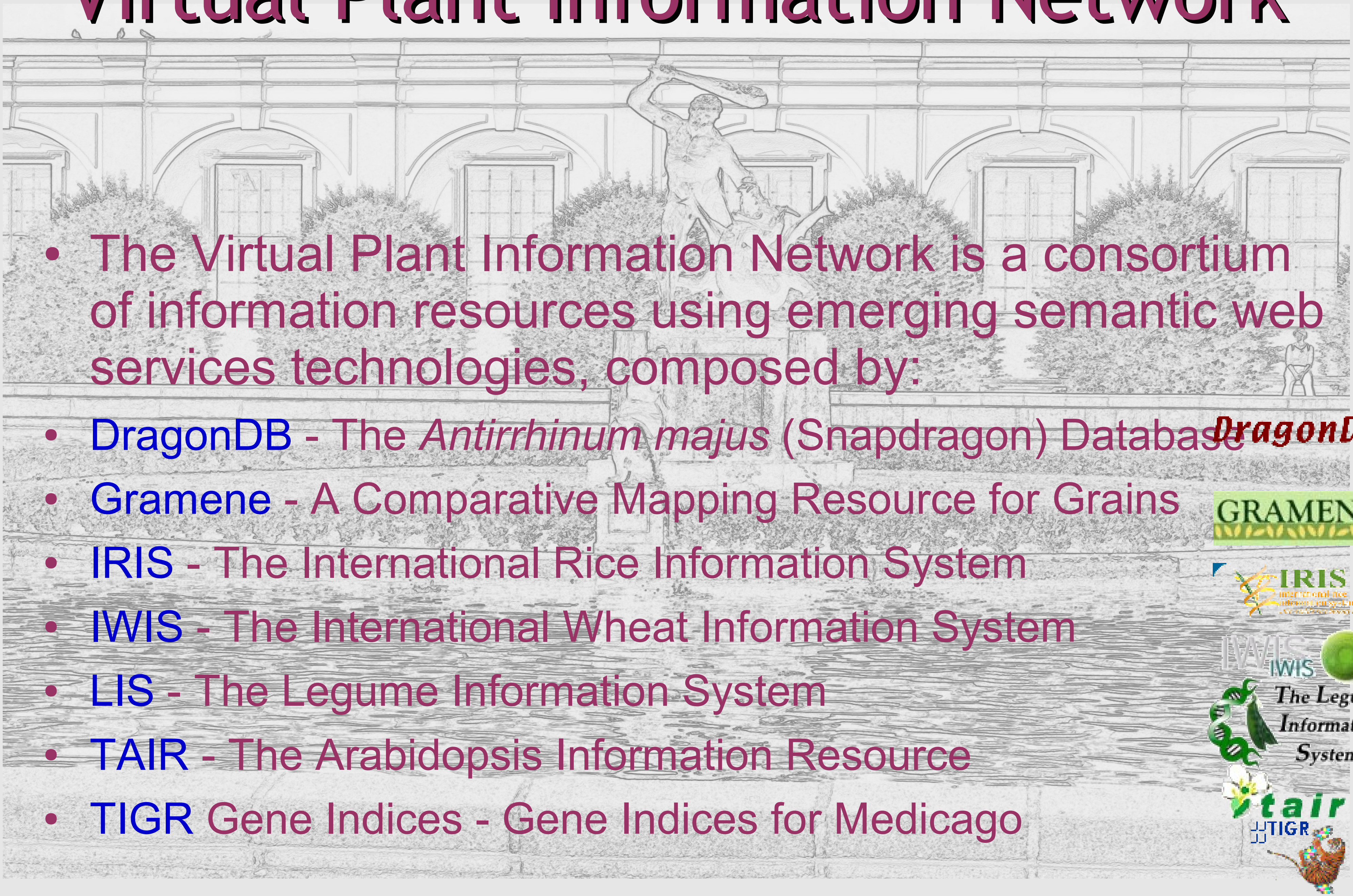
Another perspective: S-MOBY



Semantic MOBY

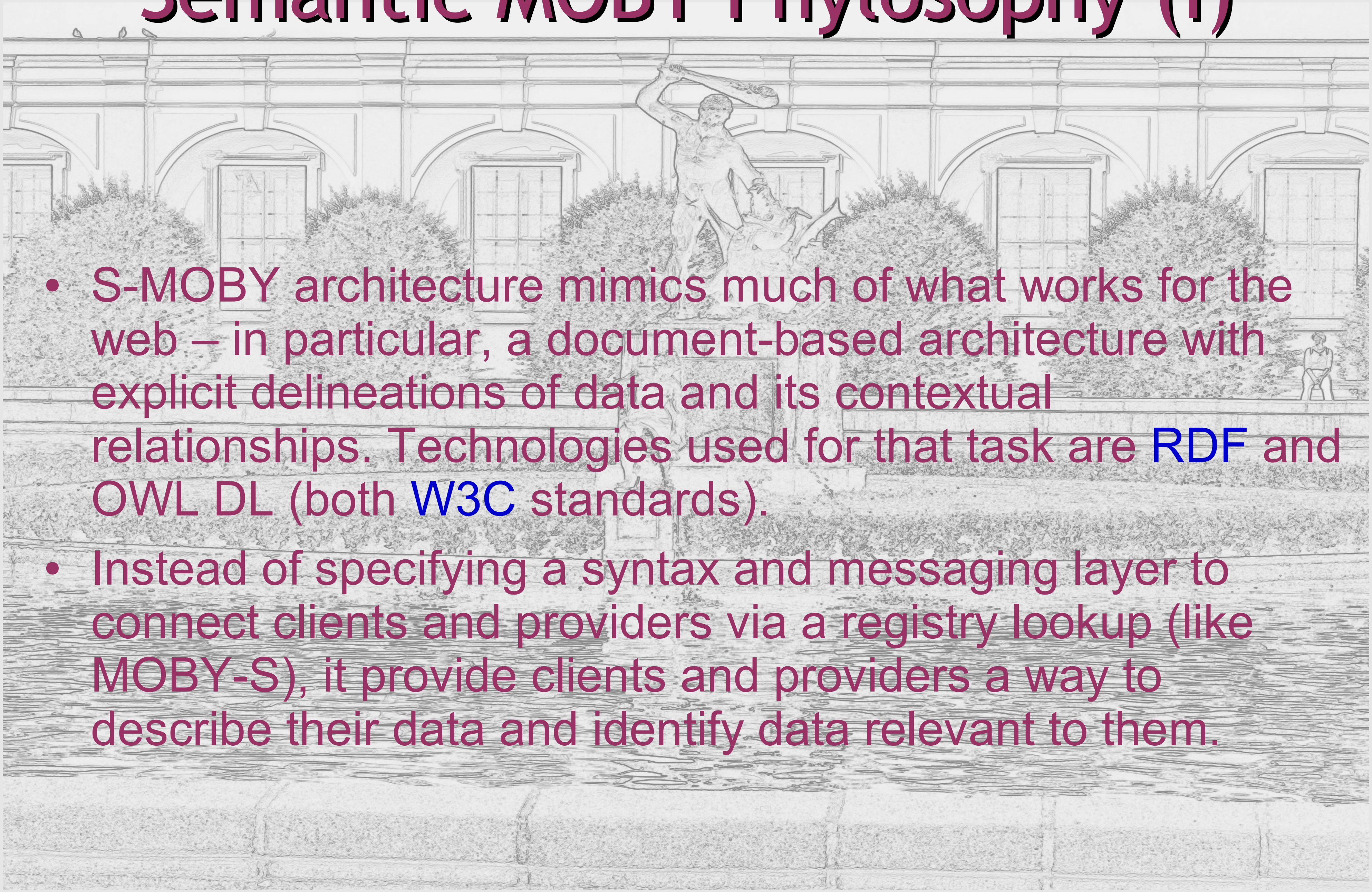
- Semantic MOBY (www.semanticmoby.org) is a Semantic Web-based architecture and open source software project for integrating web-hosted resources in the field of bioinformatics.
- Although this project was frozen more than one year, now it has been invigorated by VPIN (Virtual Plant Information Network) partners.
- Both the MOBY-S and S-MOBY projects are hoping to merge into a common framework within the next year or so.

Virtual Plant Information Network



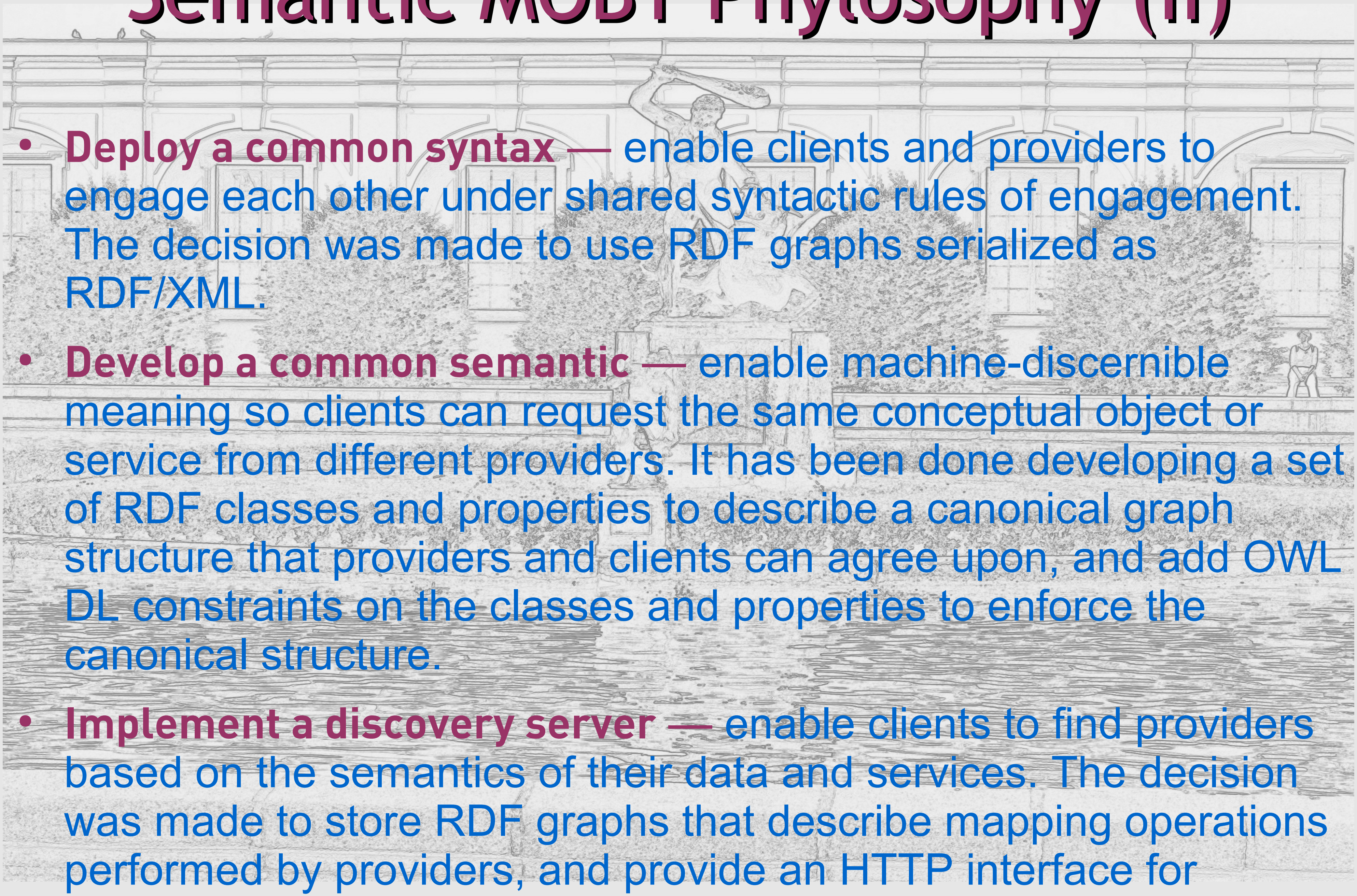
- The Virtual Plant Information Network is a consortium of information resources using emerging semantic web services technologies, composed by:
- DragonDB - The *Antirrhinum majus* (Snapdragon) Database
- Gramene - A Comparative Mapping Resource for Grains
- IRIS - The International Rice Information System
- IWIS - The International Wheat Information System
- LIS - The Legume Information System
- TAIR - The Arabidopsis Information Resource
- TIGR Gene Indices - Gene Indices for Medicago

Semantic MOBY Phylosophy (I)



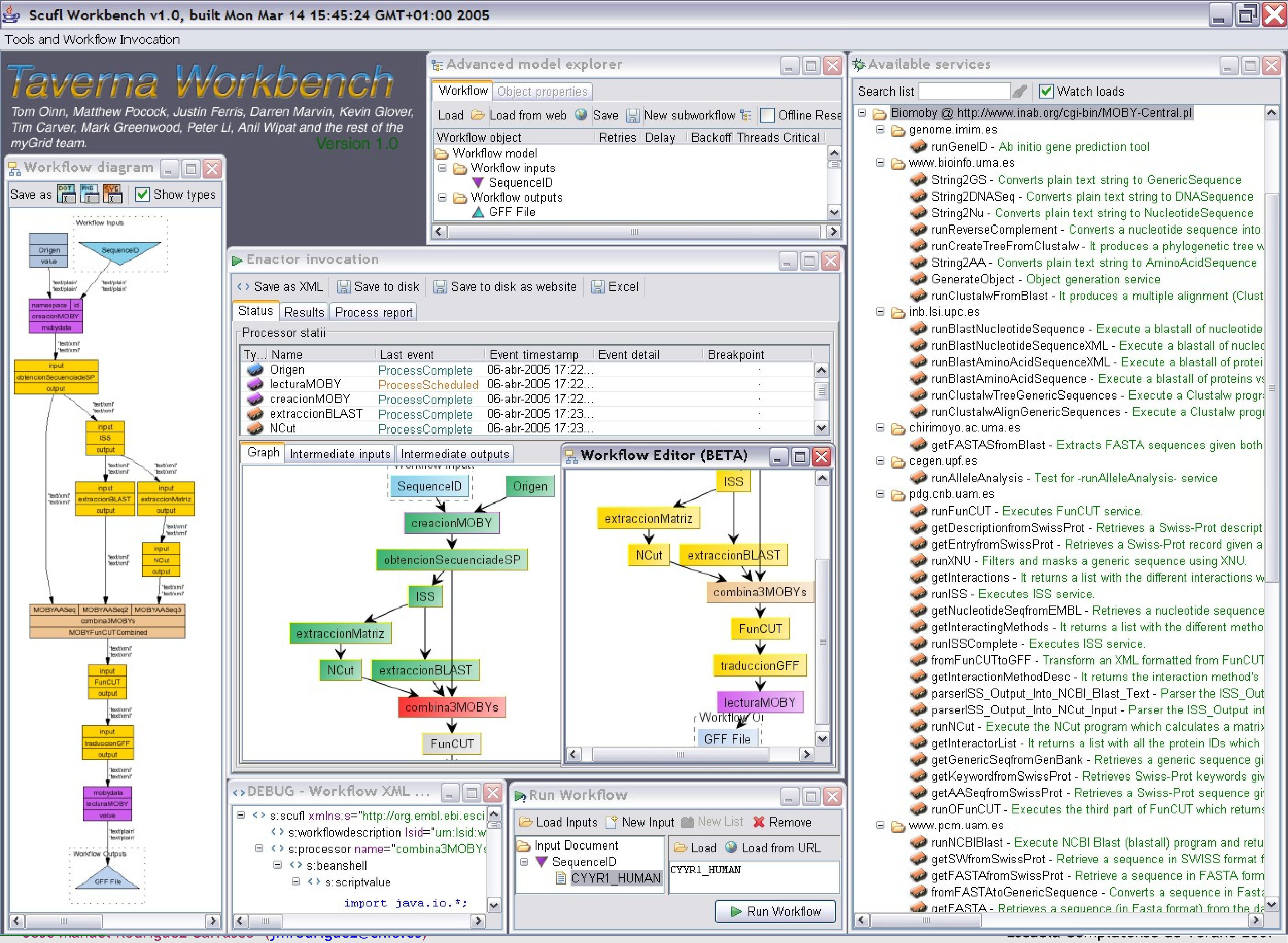
- S-MOBY architecture mimics much of what works for the web – in particular, a document-based architecture with explicit delineations of data and its contextual relationships. Technologies used for that task are RDF and OWL DL (both W3C standards).
- Instead of specifying a syntax and messaging layer to connect clients and providers via a registry lookup (like MOBY-S), it provide clients and providers a way to describe their data and identify data relevant to them.

Semantic MOBY Phylosophy (II)



- **Deploy a common syntax** — enable clients and providers to engage each other under shared syntactic rules of engagement. The decision was made to use RDF graphs serialized as RDF/XML.
- **Develop a common semantic** — enable machine-discernible meaning so clients can request the same conceptual object or service from different providers. It has been done developing a set of RDF classes and properties to describe a canonical graph structure that providers and clients can agree upon, and add OWL DL constraints on the classes and properties to enforce the canonical structure.
- **Implement a discovery server** — enable clients to find providers based on the semantics of their data and services. The decision was made to store RDF graphs that describe mapping operations performed by providers, and provide an HTTP interface for searching the repository.

José Manuel Rodríguez Carrasco (jmrodriguez@cnio.es)
José María Fernández González (jmfernandez@cnio.es)



Scufl Workbench v1.0, built Mon Mar 14 15:45:24 GMT+01:00 2005

Tools and Workflow Invocation

Taverna Workbench

Tom Oinn, Matthew Pocock, Justin Ferris, Darren Marvin, Kevin Glover, Tim Carver, Mark Greenwood, Peter Li, Anil Wipat and the rest of the myGrid team.

Version 1.0

Workflow object	Retries	Delay	Backoff	Threads	Critical
Workflow model					
Workflow inputs					
SequenceID					
Workflow outputs					
GFF File					

Processor statii

Ty...	Name	Last event	Event timestamp	Event detail	Breakpoint
	Origen	ProcessComplete	06-abr-2005 17:22...		·
	lecturaMOBY	ProcessScheduled	06-abr-2005 17:22...		·
	creacionMOBY	ProcessComplete	06-abr-2005 17:22...		·
	extraccionBLAST	ProcessComplete	06-abr-2005 17:23...		·
	NCut	ProcessComplete	06-abr-2005 17:23...		·

```
<> s:scufl xmlns:s="http://org.embl.ebi.esci
   <> s:workflowdescription lsid="urn:lsid:w
   <> s:processor name="combina3MOBYs
      <> s:beanshell
         <> s:scriptvalue
                import java.io.*;
```

Input Document
- SequenceID
  - CYYR1_HUMAN

CYYR1_HUMAN

- Biomoby @ http://www.inab.org/cgi-bin/MOBY-Central.pl
  - genome.imim.es
    - runGeneID - Ab initio gene prediction tool
  - www.bioinfo.uma.es
    - String2GS - Converts plain text string to GenericSequence
    - String2DNASeq - Converts plain text string to DNASequence
    - String2Nu - Converts plain text string to NucleotideSequence
    - runReverseComplement - Converts a nucleotide sequence into
    - runCreateTreeFromClustalw - It produces a phylogenetic tree w
    - String2AA - Converts plain text string to AminoAcidSequence
    - GenerateObject - Object generation service
    - runClustalwFromBlast - It produces a multiple alignment (Clust
  - inb.lsi.upc.es
    - runBlastNucleotideSequence - Execute a blastall of nucleotide
    - runBlastNucleotideSequenceXML - Execute a blastall of nucle
    - runBlastAminoAcidSequenceXML - Execute a blastall of protei
    - runBlastAminoAcidSequence - Execute a blastall of proteins vs
    - runClustalwTreeGenericSequences - Execute a Clustalw progr
    - runClustalwAlignGenericSequences - Execute a Clustalw progr
  - chirimoyo.ac.uma.es
    - getFASTASfromBlast - Extracts FASTA sequences given both
  - cegen.upf.es
    - runAlleleAnalysis - Test for -runAlleleAnalysis- service
  - pdg.cnb.uam.es
    - runFunCUT - Executes FunCUT service.
    - getDescriptionfromSwissProt - Retrieves a Swiss-Prot descript
    - getEntryfromSwissProt - Retrieves a Swiss-Prot record given a
    - runXNU - Filters and masks a generic sequence using XNU.
    - getInteractions - It returns a list with the different interactions w
    - runISS - Executes ISS service.
    - getNucleotideSeqfromEMBL - Retrieves a nucleotide sequence
    - getInteractingMethods - It returns a list with the different metho
    - runISSComplete - Executes ISS service.
    - fromFunCUTtoGFF - Transform an XML formatted from FunCUT
    - getInteractionMethodDesc - It returns the interaction method's
    - parserISS_Output_Into_NCBI_Blast_Text - Parser the ISS_Out
    - parserISS_Output_Into_NCut_Input - Parser the ISS_Output int
    - runNCut - Execute the NCut program which calculates a matrix
    - getInteractorList - It returns a list with all the protein IDs which
    - getGenericSeqfromGenBank - Retrieves a generic sequence gi
    - getKeywordfromSwissProt - Retrieves Swiss-Prot keywords giv
    - getAASeqfromSwissProt - Retrieves a Swiss-Prot sequence gi
    - runOFunCUT - Executes the third part of FunCUT which returns
  - www.pcm.uam.es
    - runNCBIBlast - Execute NCBI Blast (blastall) program and retu
    - getSWfromSwissProt - Retrieve a sequence in SWISS format f
    - getFASTAfromSwissProt - Retrieve a sequence in FASTA form
    - fromFASTAtoGenericSequence - Converts a sequence in Fasta
    - getFASTA - Retrieves a sequence (in Fasta format) from the da