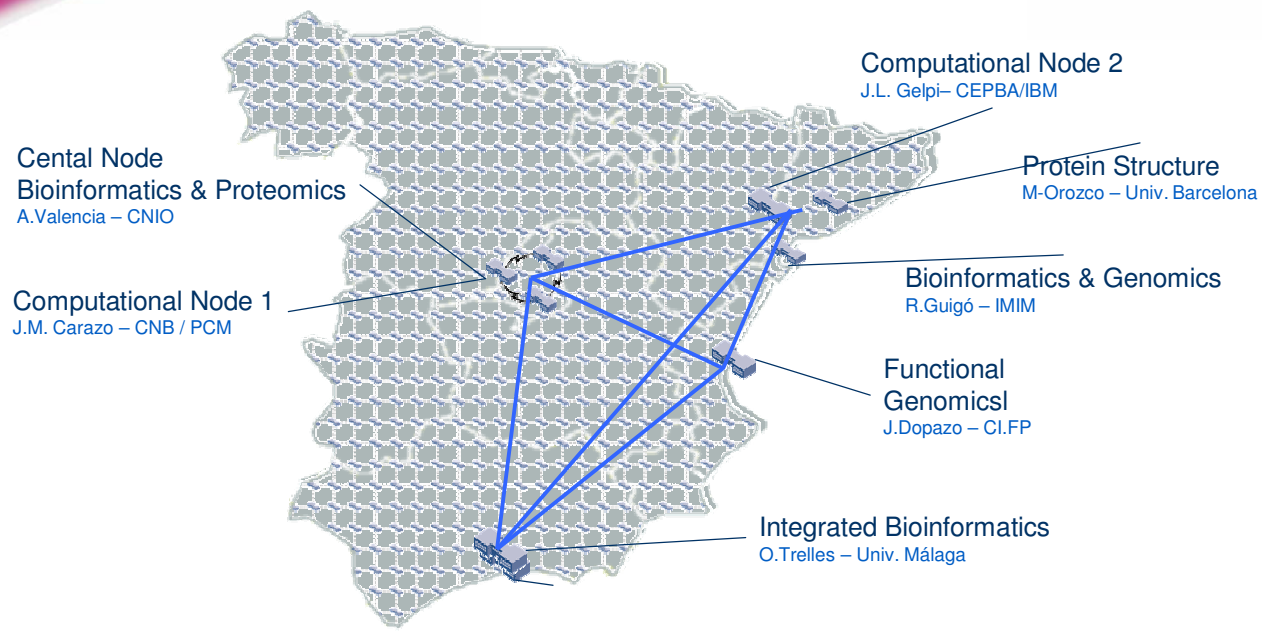
Data, Services and Computational Resources Integration at the INB

Oswaldo Trelles
University of Malaga, Spain
Integrated Bioinformatics - INB

National Institute for Bioinformatics
www.inab.org
Integrated Bioinformatics



Scalable & expandable "Virtual" Organization

The INB mission

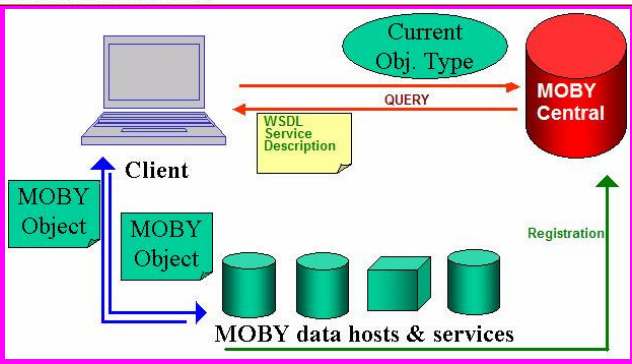
"produce and apply bioinformatic solutions in the development and workout of genomic and proteomics projects"

- Promote the **Bioinformatics dissemination** in Spain.
- Collaborate and provide **technical & scientific support** in G&P projects.
- Contribute in the **emergence and consolidation of local groups** (research or service's bioinformatics.
- Provide basic **Training in bioinformatics**.
- **Activate bioinformatics projects** from the INB.
- Promote the development and competitiveness of the **companies** in this field.
- **Internationalisation** of the INB activities.

Main lines of work at GNV5

- Integration architecture
  - MOWServ client - connecting and presenting the service offering of INB
- BioMOBY proposals
  - Protocol improvements (error handling, async, mirroring)
- Identification and restructuring of MOWServ internals
  - MOWServ is a generic client (uniform interface) but more specific clients are needed (jORCA)
- Development of software libraries

Design specifications



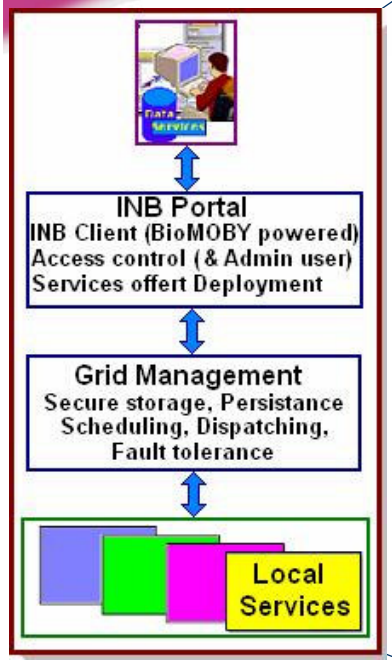
- **Persistence** of user data files
- **Controlled** Ontology and Services Registry
- Task **scheduling**, Robustness (long services)
- Objects & tasks **monitoring**
- **Generic** Client (**extended set of facilities**)
- **Creation** & visualization functionality
- Help & **training** support

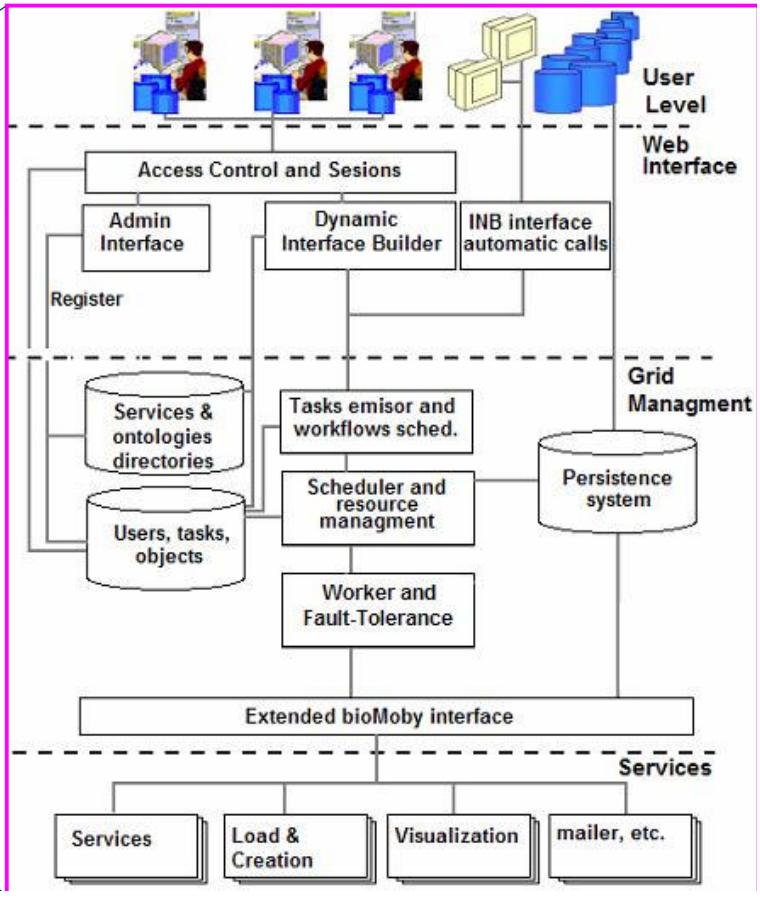
After careful analysis, the open source standard bioMOBY (Wilkinson and Links; 2002; Wilkinson, et al.; 2003) –included its pros and cons- was chosen as the underlying protocol to sustain inter-operational conjunction of nodes.

Although from the beginning we realized of some limitations in the protocol, the advantages for a rapid deployment of results as well as the increasing number of services becoming available under this platform were on top of benefits.

One of the important tasks from the INBG perspective was to extend the protocol to fulfil our objectives

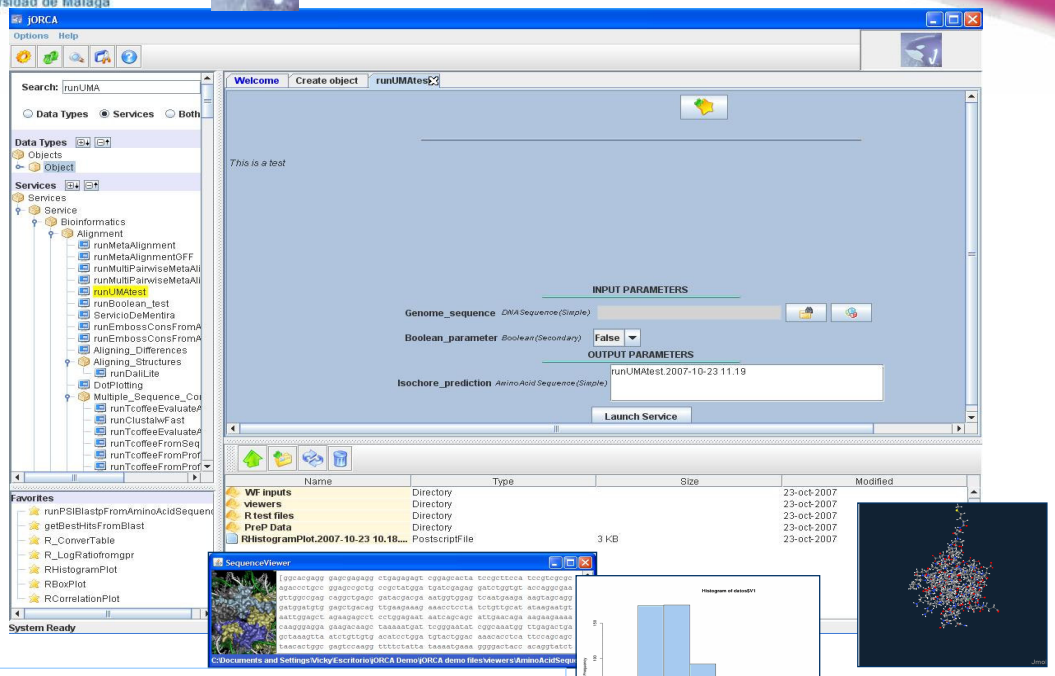
The MOWServ-v1 Architecture





jORCA -v1 : easily integration

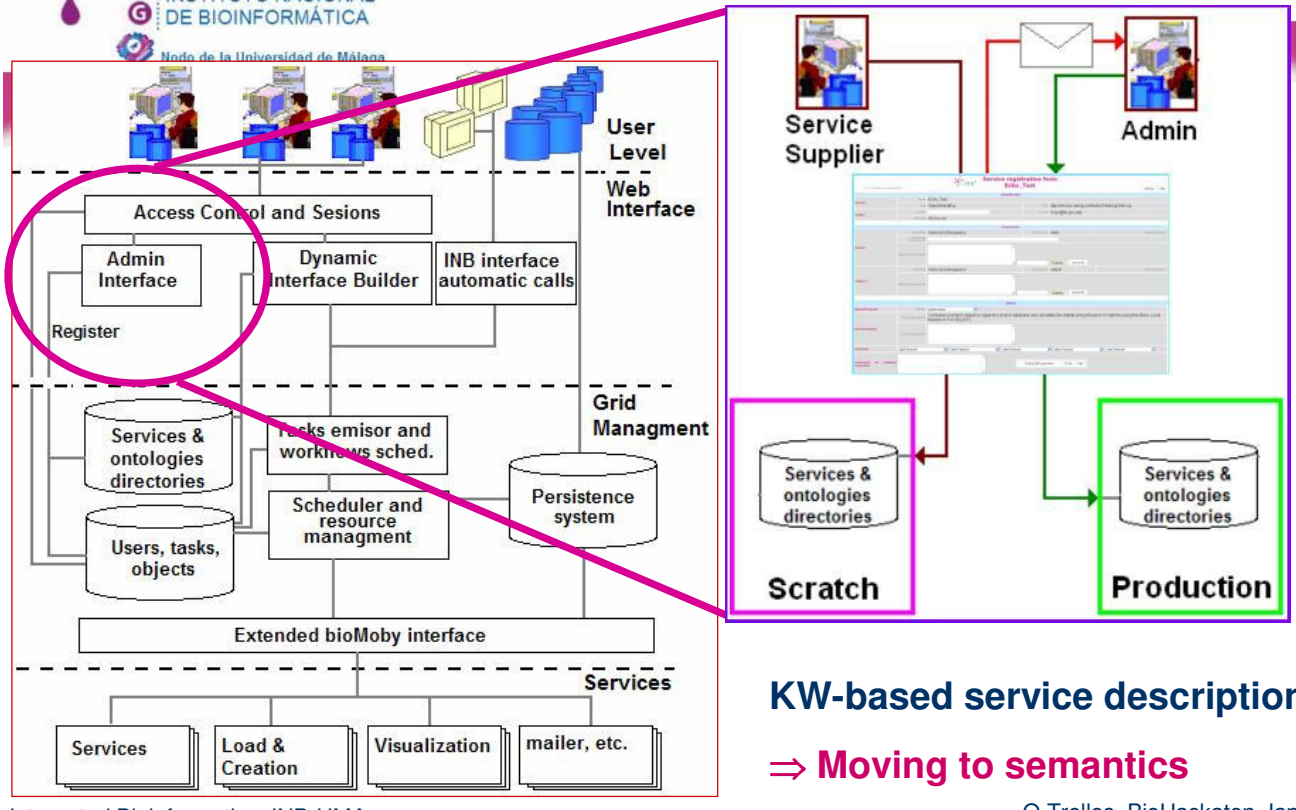
http://www.bitlab-es.com/jorca



jORCA: new Desktop interactive client: management of moby, soap, grid-based and any other type of services

Controlled Registry System

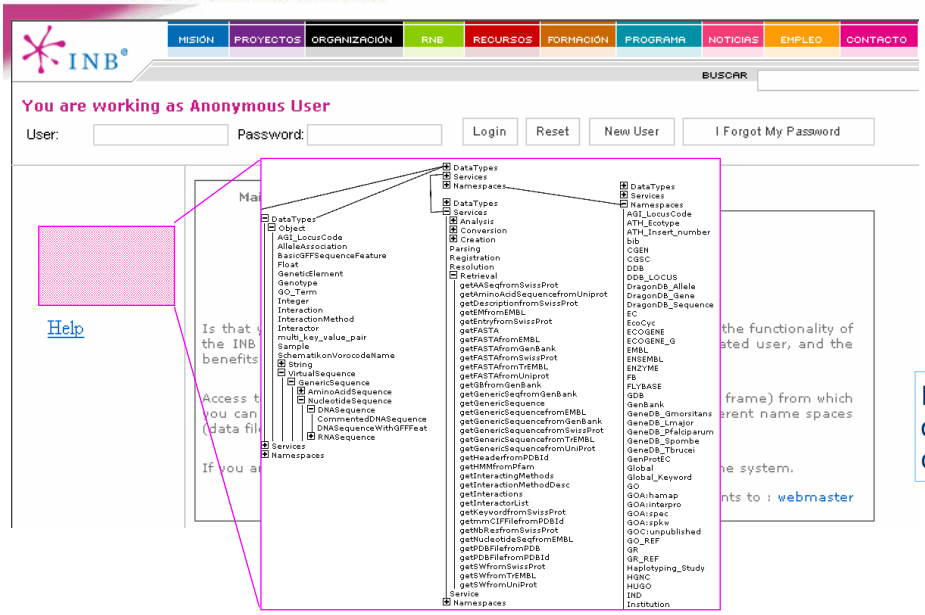
One of the most important "procedural" MOWServ's components



KW-based service description

⇒ Moving to semantics

Browsing Objects, Services & Namespaces



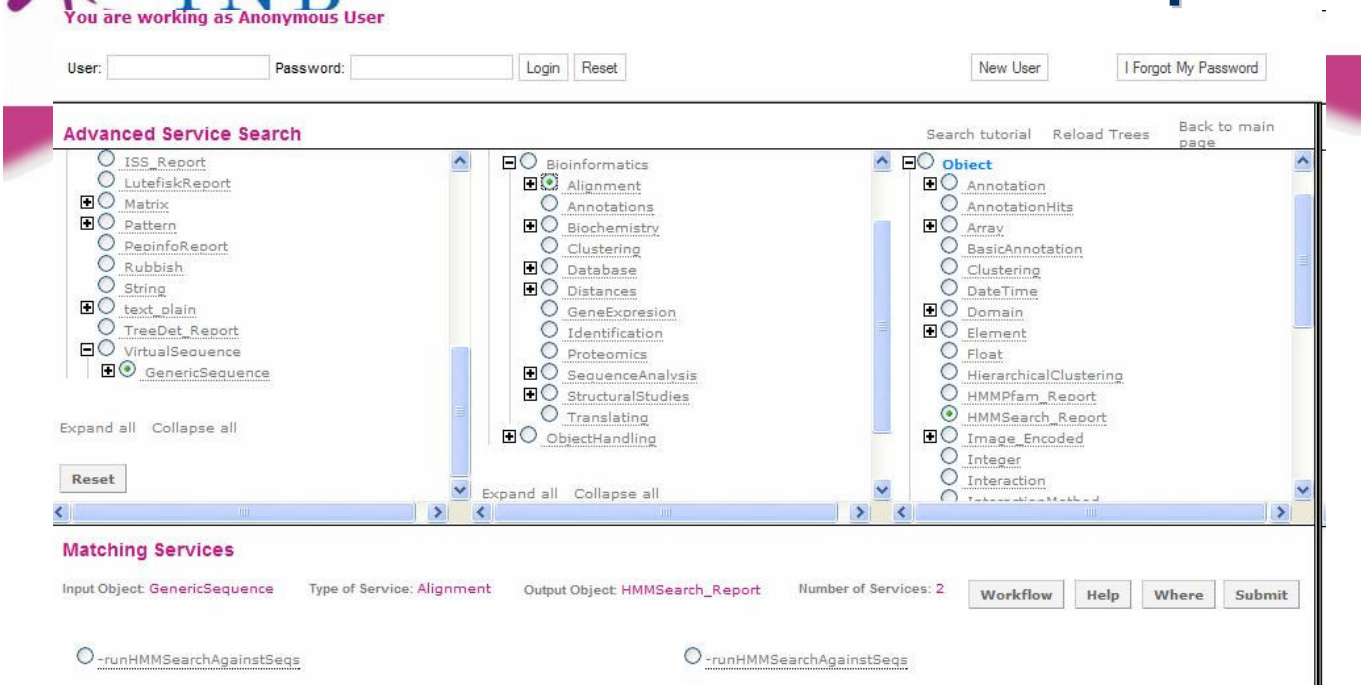
Quality control: registering / daily availability / documentation

Favourites section

Dynamic & automatic offer deployment in browseable tree organization

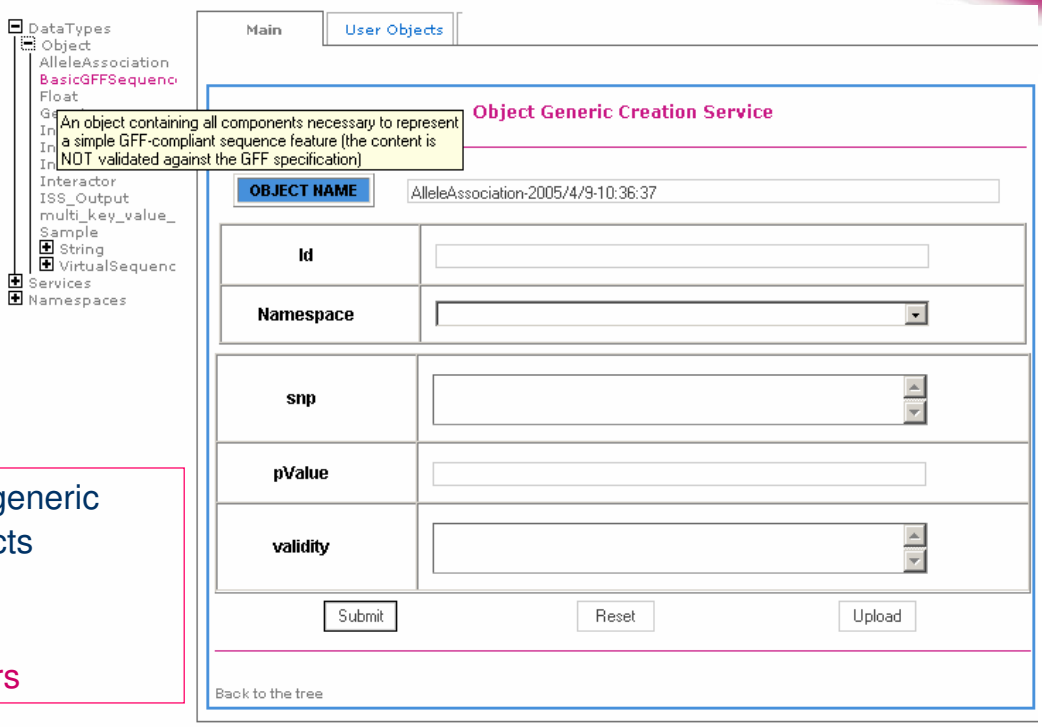
Intelligent client for integrating bioinformatics services", Navas-Delgado, et al. Bioinformatics, vol.22 no.1 2006 pages 106-111

Advanced search: Example



Example: a user has a new sequence and wants to obtain a quality alignment using HMM in order to get the family sequence where the query sequence belongs. Select GenericSequence as Input Object, and next HMMSearch_Report as output object. In this way, 4 services are suggested. Then select Alignment in Service window, and only 2 services (the same with 2 mirrors): runHMMSearchAgainstSeqs, are identified which is useful for the final target.

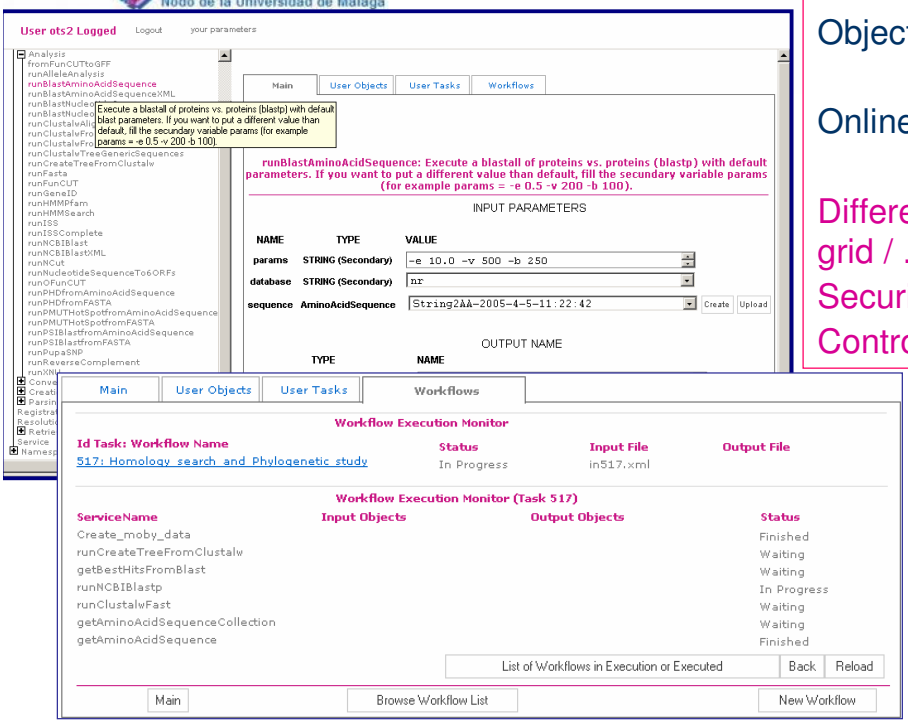
Creation Services



Creation services: generic
Up-down load objects

Plug-in specific
Transparent to users

Service invocation

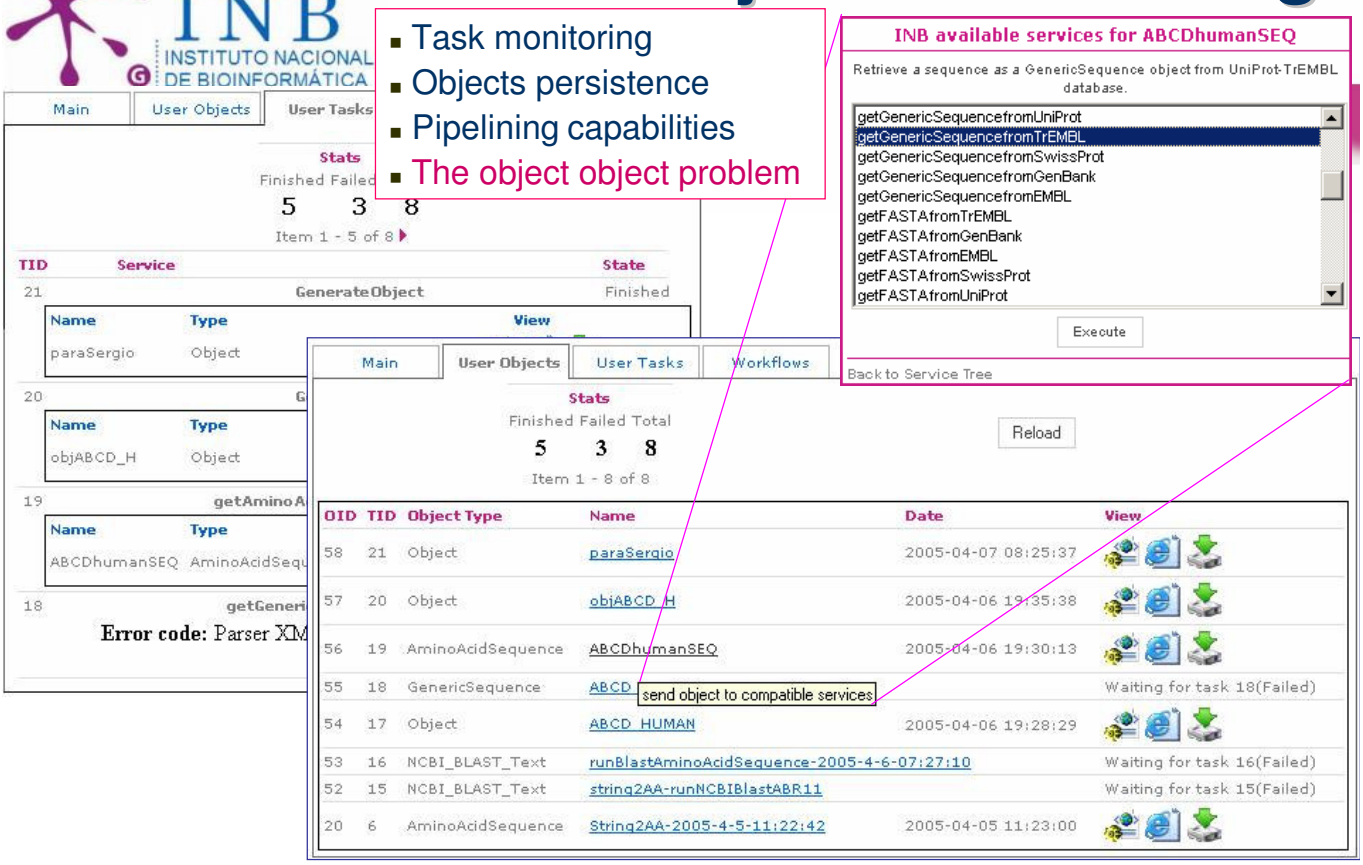


Automatic / Uniform
Objects: online creation
Up-download
Online help services

Different protocols (soap / Moby / grid / ... jORCA: plug-in workers)
Security issues (anonymous data)
Controlled access to resources

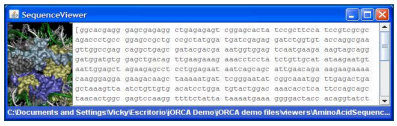
Tasks & Objects Monitoring

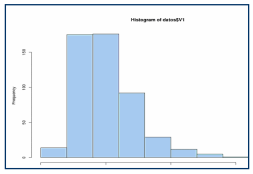
- Task monitoring
- Objects persistence
- Pipelining capabilities
- The object object problem

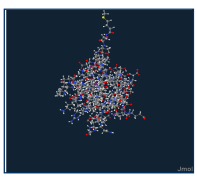


Viewing results

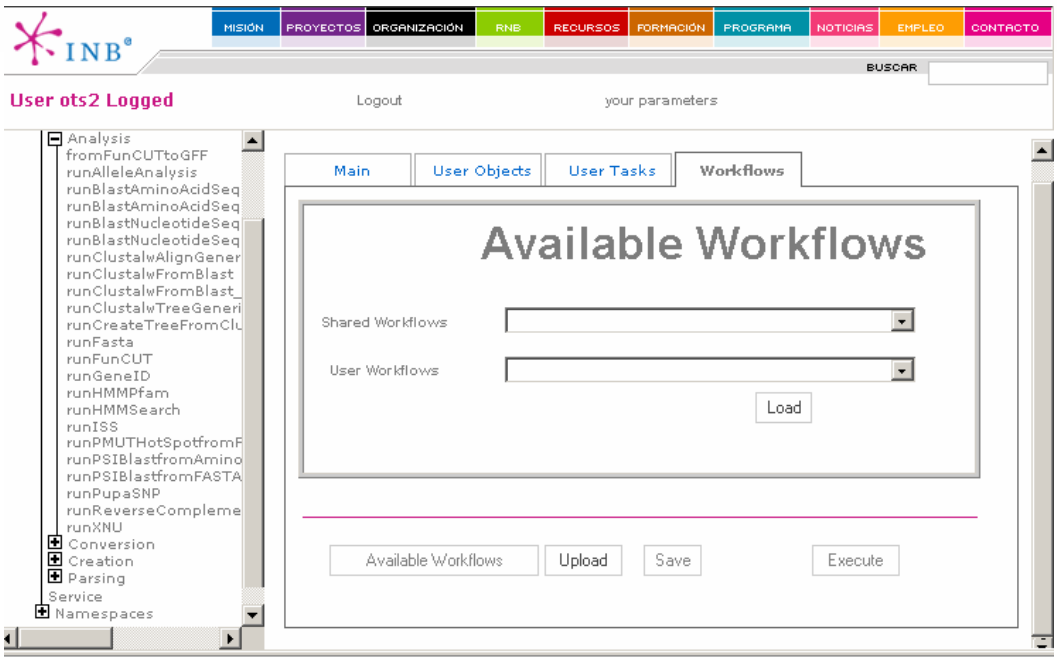
- Plug-in based architecture:
  - Each object can be associated with a XSLT file
- Service providers can submit specialized viewers for their data (e.g. jMOL)
- Results always viewable as XML or HTML.
- Improvements: “Interactive” visualization / registering





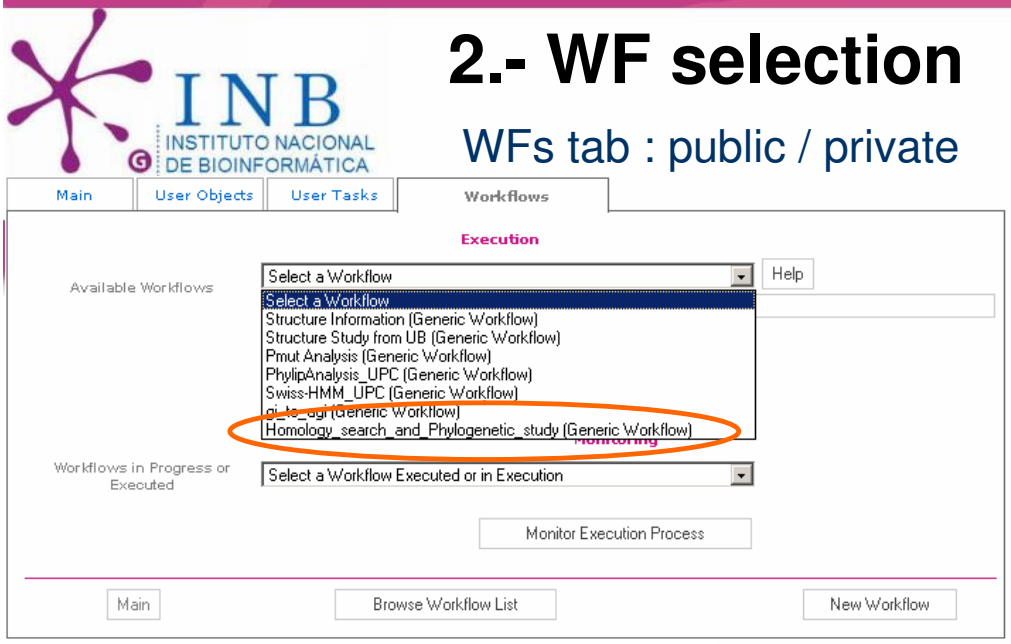


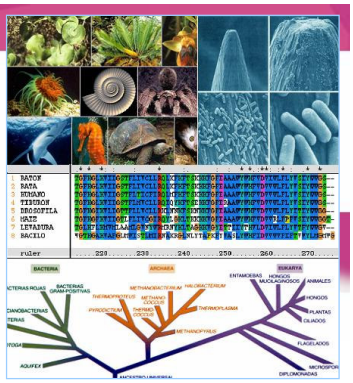
Workflows



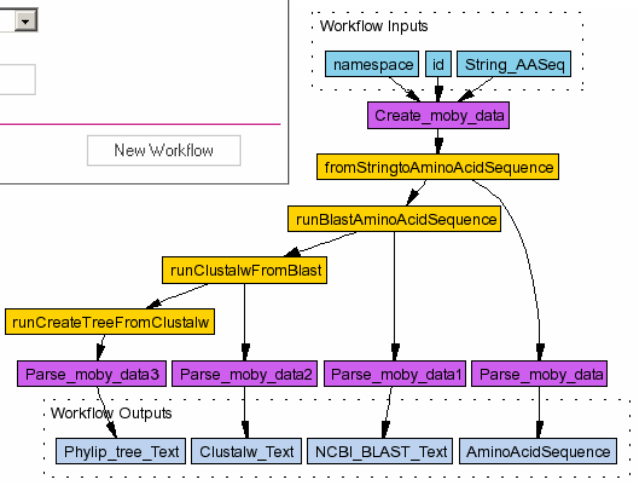
2.- WF selection

WFs tab : public / private





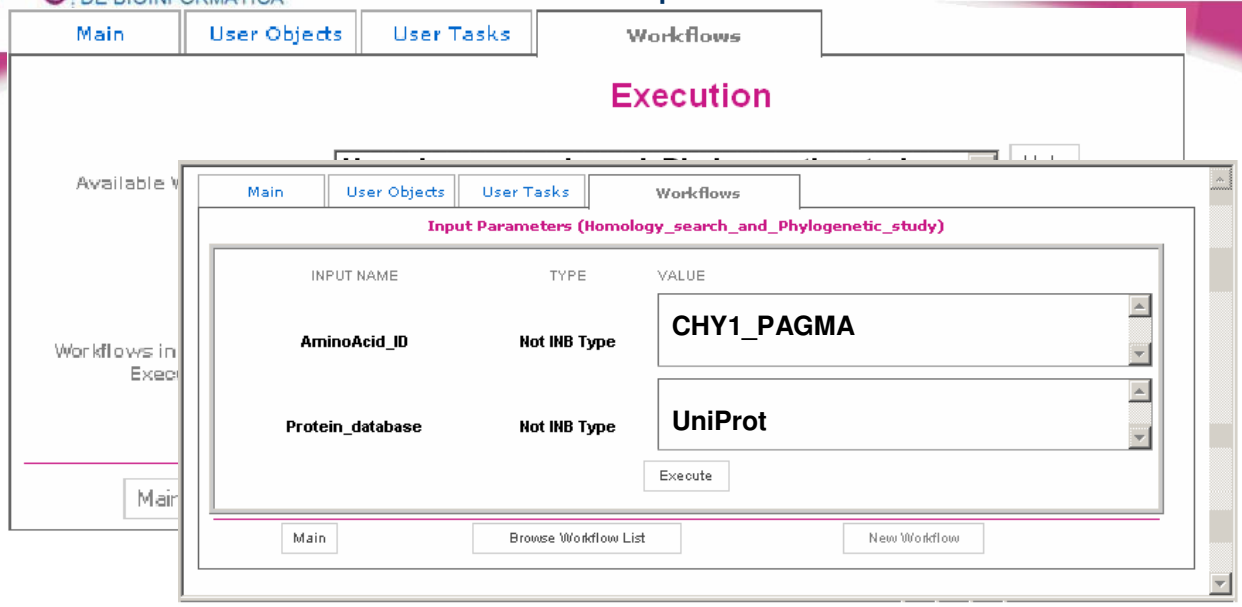
Phylogenetic tree



STEP 1: Sequence retrieval
STEP 2: Blast
STEP 3: Best hits
STEP 2: Multiple seq. alignment
STEP 3: Phylogenetic tree

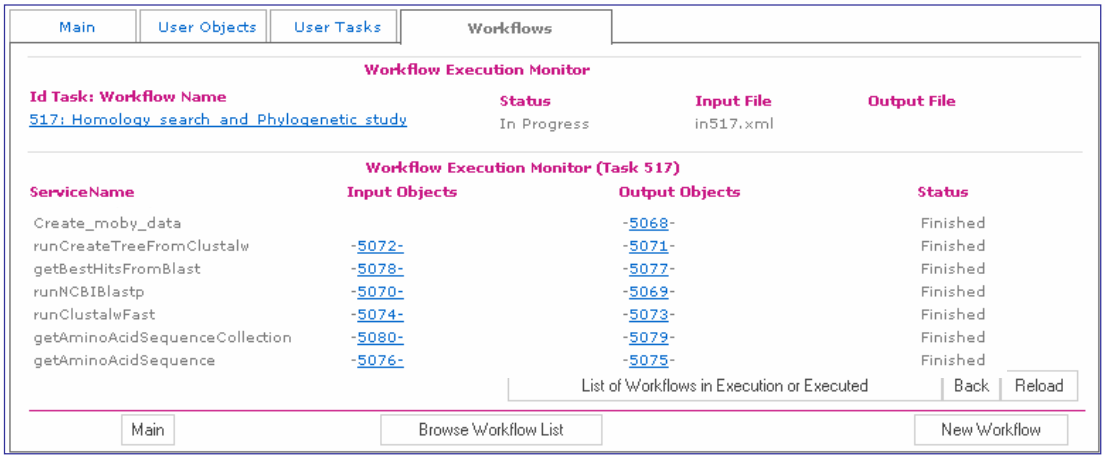
3.- specialise

Input definition



Launch execution

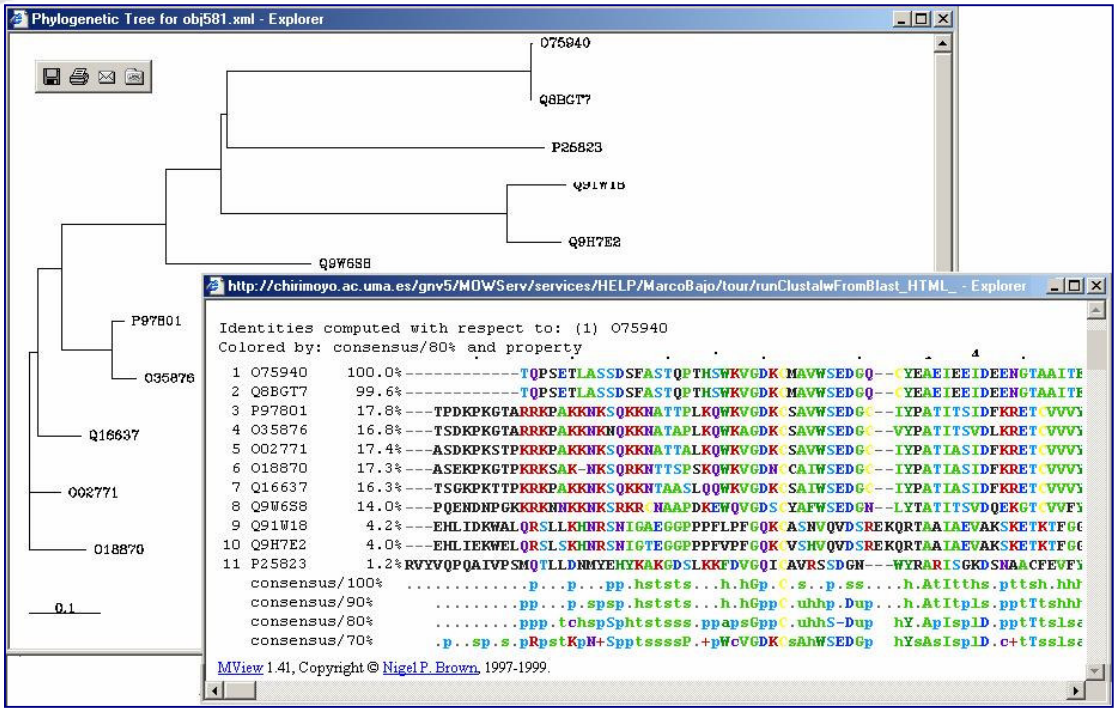
4.- Monitoring



In progress ....

...done

5.- Results



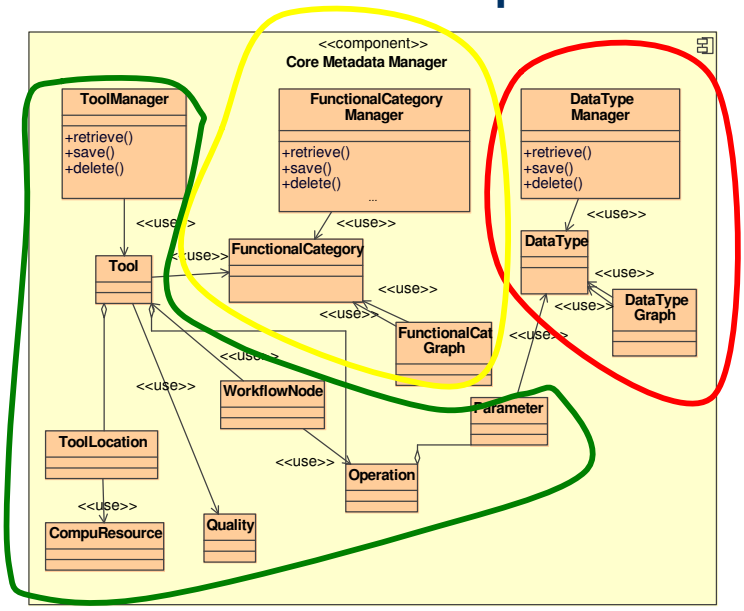
Open questions

Short term

- Asynchronous services (& progress status notification)
- Error handling
- Replication (mirroring) system
- Daily Service quality control
- Service's help system
- Advanced service discovering
- Collections
- Large data sets (network overloading)
- Indirect (by reference) pass of data (locality of D & S)
- Workflows (storage in the repository, WWE-jmf)
- Semantic annotations (datatypes & Services)
- Define services categories
- New datatypes: Gene expression; images; ...

Open questions
long term

- Review / redefine the repository data model
  - modular / distributed / users / persistence / security...



Integrated Bioinformatics Node

www.bitlab-es.com

Bioinformatics and Information Technologies Laboratory
http://chirimoyo.ac.uma.es/bitlab

Arquitectura de Computadores

Lenguajes y Ciencias de la Computación

Genética

Biología Molecular y Bioquímica