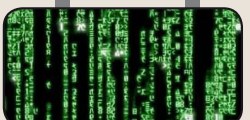
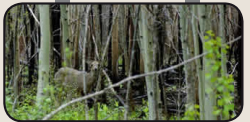
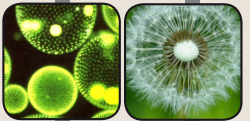
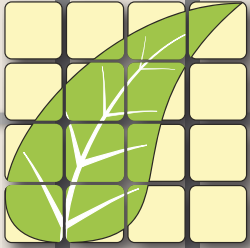


BioVeL

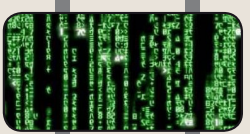
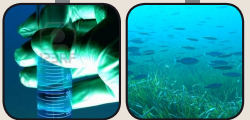
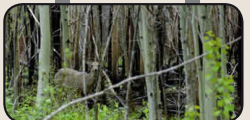
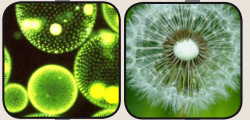


Outlook

- Overview of the BioVeL project
- Overview of the workflow framework
- Features supported:
 - Application Management
 - Data Management
- Frontend and Backend overview
- First test and results
- Conclusions and Work in progress



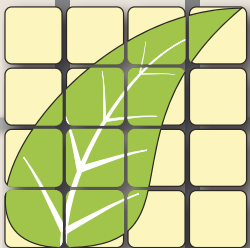
BioVeL



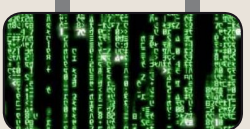
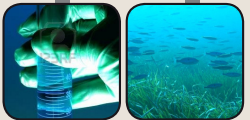
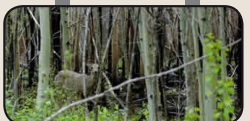
Biodiversity Virtual e-Laboratory

BioVeL is an international network of experts

- Connects two scientific communities: IT and biodiversity.
- Offers an international network of IT expert scientists in BioVeL's data processing services.
- Shares expertise in workflow studies among BioVeL's users.
- Fosters an international community of researchers and partners on biodiversity issues.
- BioVeL is an e-laboratory that supports research on biodiversity using large amounts of data from cross-disciplinary sources.



BioVeL



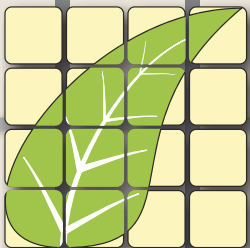
Biodiversity Virtual e-Laboratory

BioVeL is a consortium of 15 partners from 9 countries

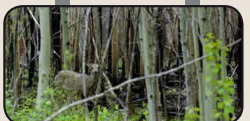
1. Cardiff University, UK – Coordinator
2. Centro de Referência em Informação Ambiental, Brazil
3. Foundation for Research on Biodiversity, France
4. Fraunhofer-Gesellschaft, Institute IAIS, Germany
5. Free University of Berlin – Botanical Gardens and Botanical Museum, Germany
6. Hungarian Academy of Sciences Institute of Ecology and Botany, Hungary
7. Max Planck Society, MPI for Marine Microbiology, Germany
8. National Institute of Nuclear Physics, Italy
9. National Research Council: Institute for Biomedical Technologies and Institute of Biomembrane and Bioenergetics, Italy
10. Netherlands Centre for Biodiversity (NCB Naturalis), The Netherlands
11. Stichting European Grid Initiative, The Netherlands
12. University of Amsterdam, Institute of Biodiversity and Ecosystem Dynamics, The Netherlands



13. University of Eastern Finland, Finland
14. University of Gothenburg, Sweden
15. University of Manchester, UK



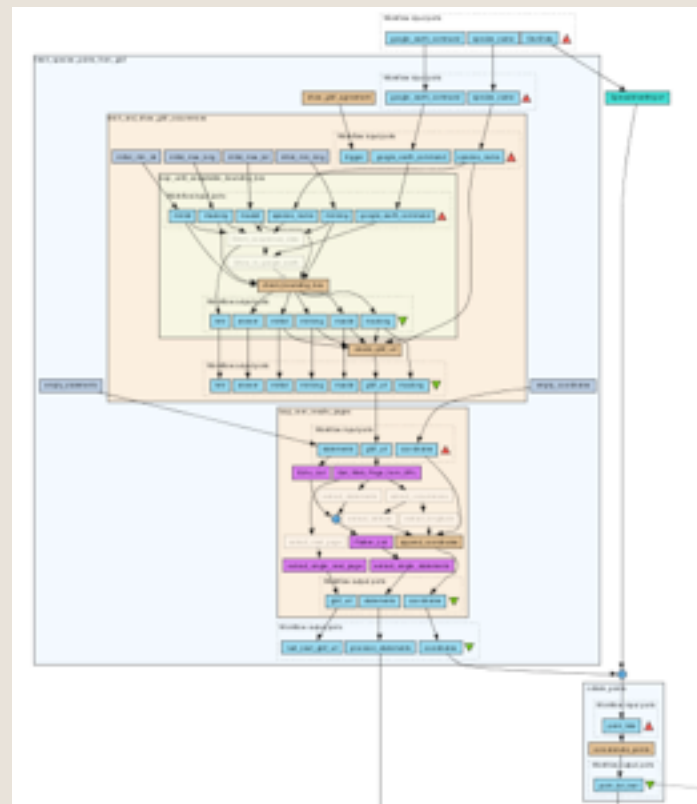
BioVeL



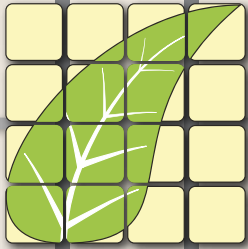
Biodiversity Virtual e-Laboratory

BioVeL is a powerful data processing tool

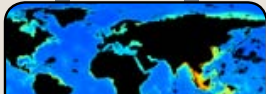
- Import data from one's own research and/or from existing libraries.
- “Workflows” (series of data analysis steps) allow to process vast amounts of data.
- Build your own workflow: select and apply successive “services” (data processing techniques.)
- Access a library of workflows and re-use existing workflows.
- Cut down research time and overhead expenses.
- Contribute to LifeWatch and GEO BON.



Part of a workflow to study the ecological niche of the horseshoe crab



BioVeL

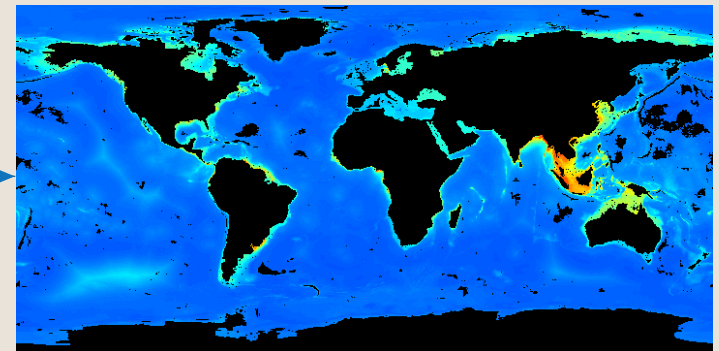
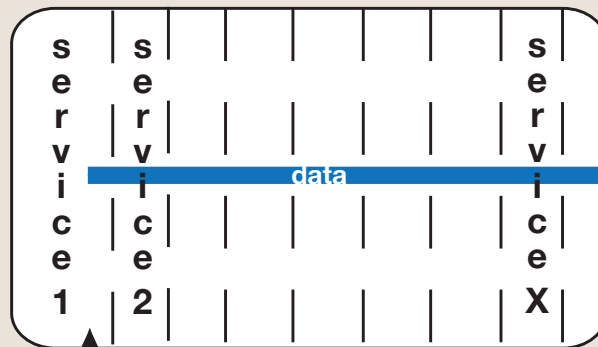


Biodiversity Virtual e-Laboratory

Showcase study 1: create a workflow*

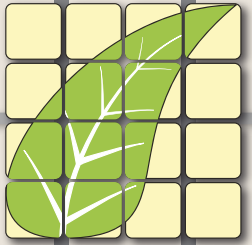
Study on the ecological niche of the south east Asian horseshoe crab, an endangered species:

- Import south east Asian data from external library
- Apply succession of “services” = workflow
- Result: ecological niche map



results: map showing the potential ecological niche of the south-east Asian horseshoe crab

* courtesy Matthias Obst, University of Gothenburg, Sweden



BioVeL



Biodiversity Virtual e-Laboratory

Showcase study 2: re-use a workflow

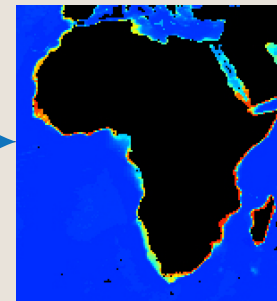
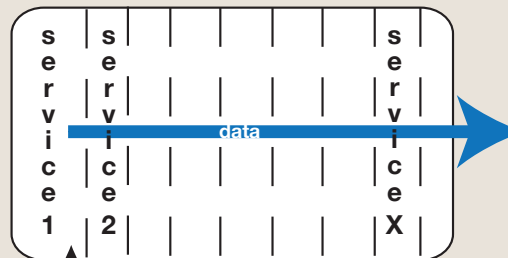
Study on the ecological niche of the American horseshoe crab

- Import American data
- Re-use south east Asia crab study workflow
- Result: ecological niche map for American horseshoe crab

Compare the ecological niches of the south east Asian and American crabs.

Potential study of the ecological niche of an African animal

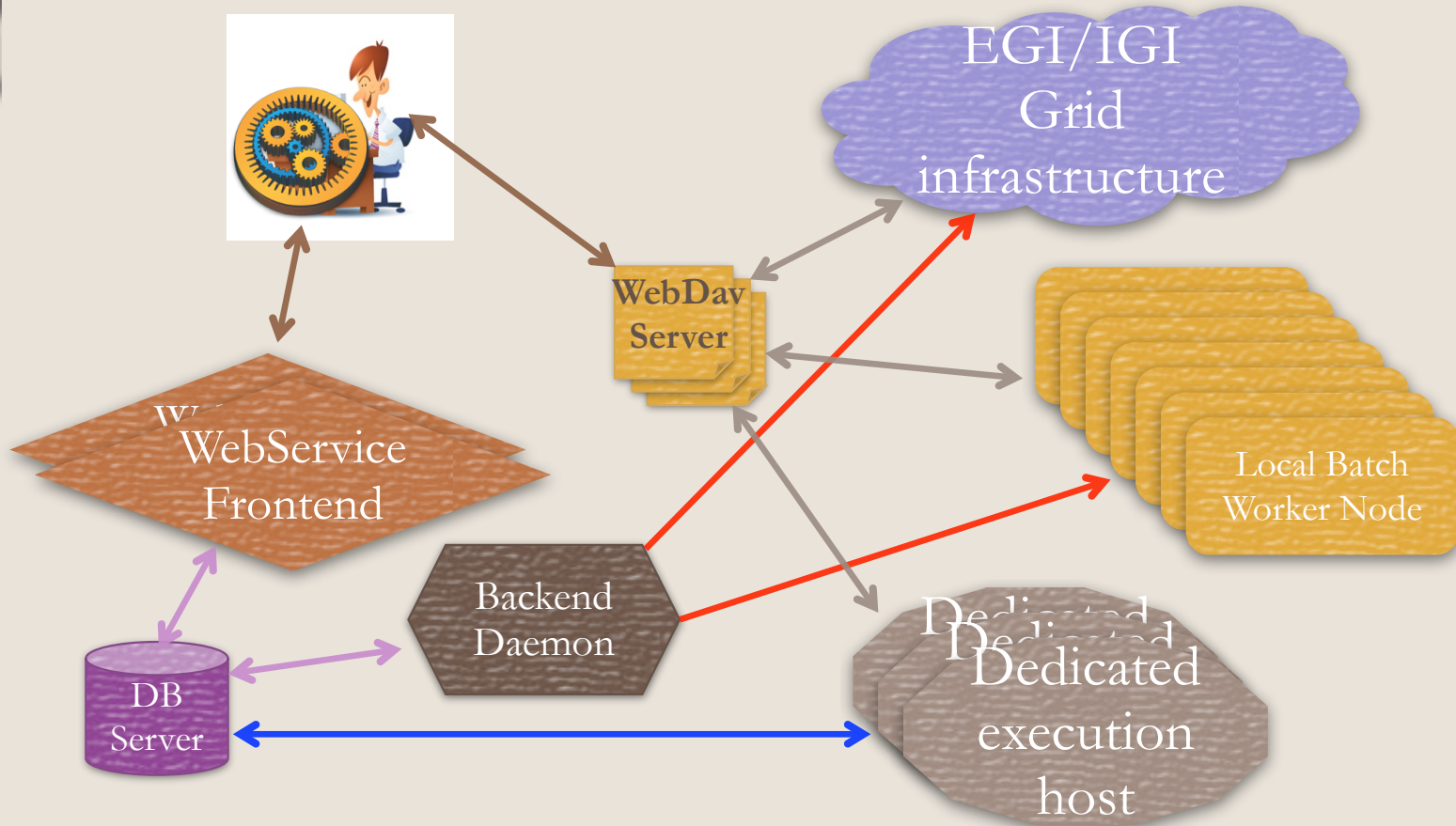
- Import African data
- Re-use horseshoe crab study workflow
- Result: ecological niche map for African animal



results: map showing the potential ecological niche of an African species re-using the same workflow --or an altered version of it-- as for the south east Asia horseshoe crab.



Web service Logical Design





General overview of the framework

- REST-FUL Web service for the FRONTEND
 - Web Server: Apache TomCat
 - DBMS: MySql 5
 - Framework Jersey
 - Framework Java EE 6.0 SDK
 - Asynchronous operations
- BACKEND written in JAVA
 - Multithread
 - Reads DB, submits and executes jobs
 - At the moment we support: PBS, EGI/IGI grid infrastructure, dedicated servers



General overview of the framework

- Each call to the web service is intended to ask for an execution of a well specified application:
 - Only supported applications (and well known to the service provider) could be executed
 - Supporting a new application is usually few days of works from the service provider point of view
 - Most of the application only requires one or few input files
 - The user can request a run, by choosing the name of the application and the name (and location) of the input files
 - You can also use a external file available through http, ftp, etc.
 - When needed the user could change also parameters used in the command line
- The output of the runs at the end will be available (also to other services) via http link



Describing the application

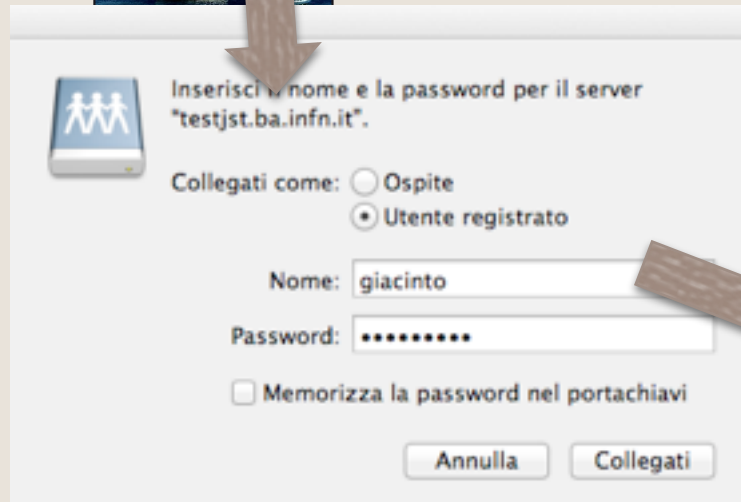
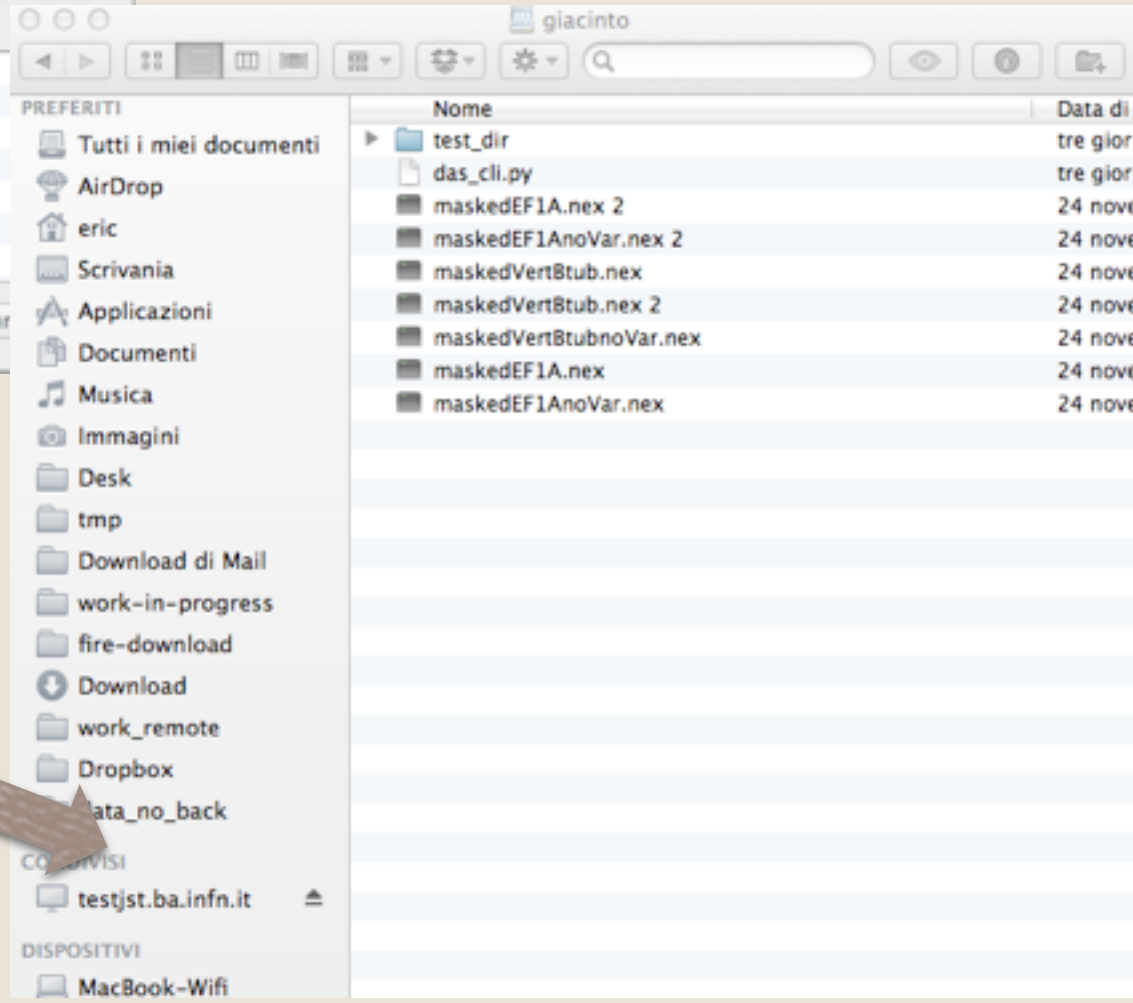
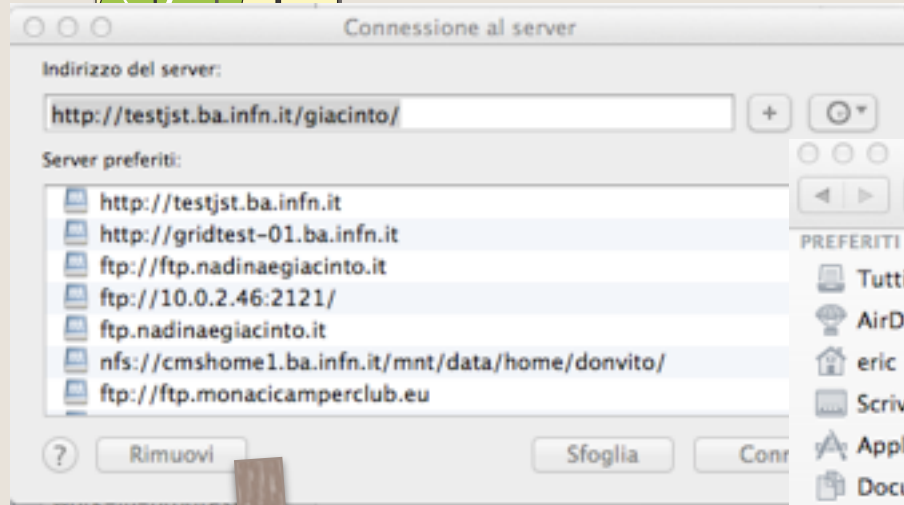
- Each application is described by:
 - A bash script that prepare the environment and run the real application
 - Hidden to the final user
 - A set parameters
 - Input location and file name
 - Arguments for the executable
- Returns:
 - Status
 - Output URL

Features Supported



- Transferring files and directory:
 - user friendly
 - good performance
 - Well established protocol (WebDav)
- Submitting and monitoring the status of one or several runs with simple REST APIs
 - Multiple submission (up to thousands tasks) with a single REST call
 - Monitoring the whole run with a single REST call

Screenshots: WebDav DataManagement Service



Screenshots:

WebDav DataManagement Service



Name	Last modified	Size	Description
Parent Directory	-	-	-
das_cli.py	26-Nov-2011 11:01	97	
maskedEF1A.nex	24-Nov-2011 15:04	89K	
maskedEF1A.nex 2	24-Nov-2011 15:04	89K	
maskedEF1AnoVar.nex	24-Nov-2011 15:04	89K	
maskedEF1AnoVar.nex 2	24-Nov-2011 15:04	89K	
maskedVertBtub.nex	24-Nov-2011 15:04	180K	
maskedVertBtub.nex 2	24-Nov-2011 15:04	180K	
maskedVertBtubnoVar.nex	24-Nov-2011 15:04	180K	
test_dir/	26-Nov-2011 11:04	-	

Apache/2.2.16 (Debian) Server at testjst.ba.infn.it Port 80

You can access those files using web browser:

You can easily share your data with others colleagues

Or use the input/output within other (web) services

Web service example

Insert Jobs:

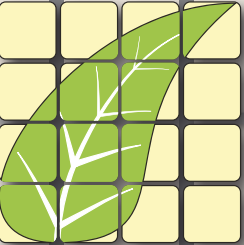
"http://localhost:8080/RestService/services/QueryJob/InsertJobs?
NAME={blast}&arguments={http://webtest.ba.infn.it/vicario/
FinalFusariumDB_2.nex ArgOne; http://webtest.ba.infn.it/vicario/
FinalFusariumDB_1.nex ArgTwo;}"

```
▼<Job>
  <Name>blast</Name>
  <Flag>f1966c8a-e926-4d17-a08a-4f83654d57ce</Flag>
  ▼<JobsID>
    <JobId>453112</JobId>
    <JobId>453113</JobId>
  </JobsID>
</Job>
```

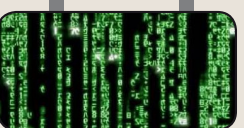
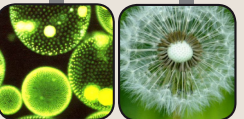
Select Jobs:

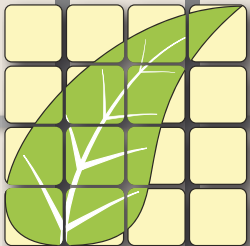
"http://localhost:8080/
RestService/services/QueryJob/
SelectJobs?FLAG={001}"

```
▼<Jobs>
  ▼<Job>
    <Arguments>1</Arguments>
    <Comment>grid</Comment>
    <CPUs>1</CPUs>
    <Flag>f1966c8a-e926-4d17-a08a-4f83654d57ce</Flag>
    <Id>453112</Id>
    <LastCheck>2011-11-28 10:30:17.0</LastCheck>
    <Name>blast</Name>
    <Output/>
    <Provenance/>
    <Status>free</Status>
  </Job>
  ▼<Job>
    <Arguments>2</Arguments>
    <Comment>grid</Comment>
    <CPUs>1</CPUs>
    <Flag>f1966c8a-e926-4d17-a08a-4f83654d57ce</Flag>
    <Id>453113</Id>
    <LastCheck>2011-11-28 10:30:17.0</LastCheck>
    <Name>blast</Name>
    <Output/>
    <Provenance/>
    <Status>free</Status>
  </Job>
</Jobs>
```

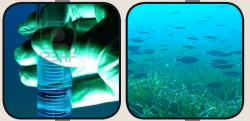
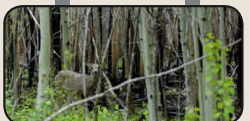


BioVeL





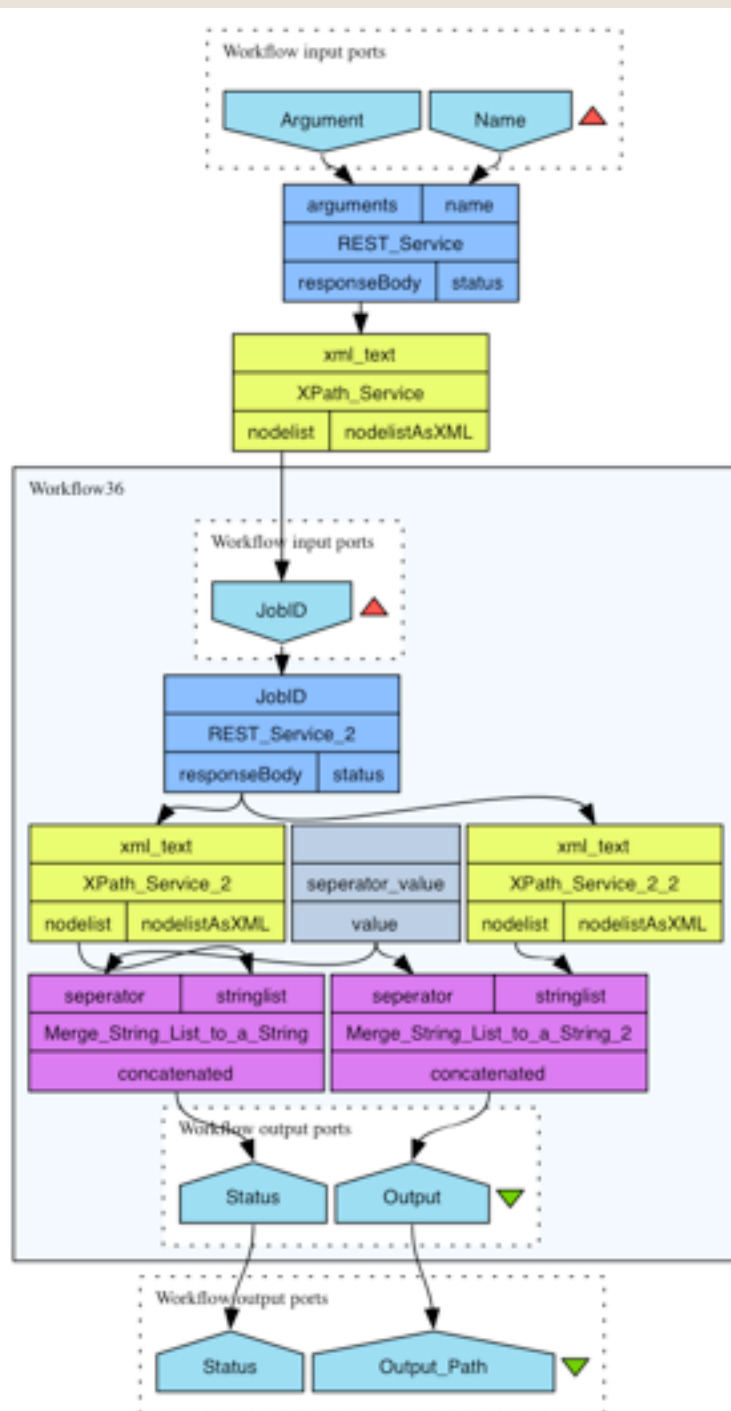
BioVeL



A test "Taverna WF" already available!!



We will provide a dedicated WF for each application supported





Description of the resources

- Grid distributed computing infrastructures (EGI Production Grid)
 - ~ 200'000 Cores, Hundreds of distributed sites
 - ~ 20PB
 - There sites that offers a good support to biomed VO
 - **Very useful for huge number of independent tasks**
- Local batch INFN-Bari farm
 - ~3500 Cores
 - 1.2PB of storage
 - Supporting several VOs
 - 10% of the share are dedicated to opportunistic VOs (like bioinformatics)
 - **From tens to hundreds of concurrent execution, good support also for MPI application**
- Dedicated servers
 - Big servers: 24Cores, 48GB of RAM (will be higher in the future)
 - Specialized servers: 2 Tesla C2070 GPUs
 - **Thread based parallel applications, GPU enabled applications, short and high priority tasks**

Test & Results



- Stress test already passed:
 - 100'000 insert in a loop... no memory leak or similar problems
 - Up to 100 concurrent clients without problems
 - 1000 tasks insert in a single REST call
 - ~1M of tasks managed from DB+backend
- A lot of experience in porting Bioinformatics application over EGI distributed computing infrastructure:
 - Hmmer, MrBayes, Blast, PAML, MUSCLE, EMBOSS, Biopython, AmpliconNoise, ABCtool, Bowtie, BayeSSC, etc...



Conclusions & To-do

- We have a high scalable and solid REST service that could be used to supports execution of applications over different computing infrastructure
- We have also a high performance data transfer and sharing service
- We will publish soon both service and WorkFlows on BioCatalogue and MyExperiments
- It will be quite easy to add new application in the near future

People involved



- Giacinto Donvito (INFN)
- Pasquale Notarangelo (INFN)
- Saverio Vicario (CNR)



Biodiversity Virtual e-Laboratory

BioVeL is funded by the European Commission 7th Framework Programme (FP7). It is part of its e-Infrastructures activity.

Under FP7, the e-Infrastructures activity is part of the Research Infrastructures programme, funded under the FP7 'Capacities' Specific Programme. It focuses on the further development and evolution of the high-capacity and high-performance communication network (GÉANT), distributed computing infrastructures (grids and clouds), supercomputer infrastructures, simulation software, scientific data infrastructures, e-Science services as well as on the adoption of e-Infrastructures by user communities.

BioVeL is free and available via internet.



www.biovel.eu, contact Alex Hardisty: HardistyAR@cardiff.ac.uk