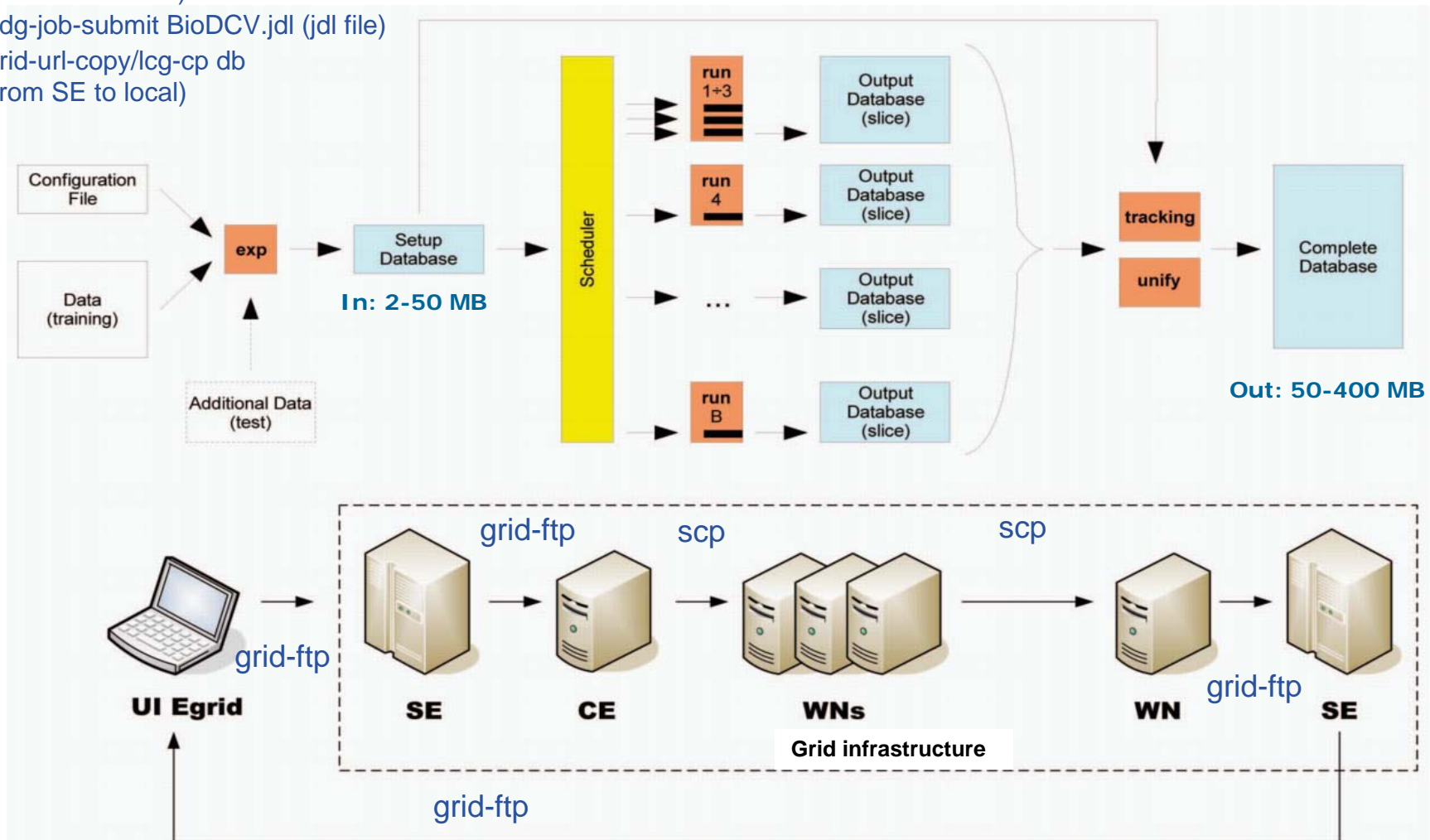


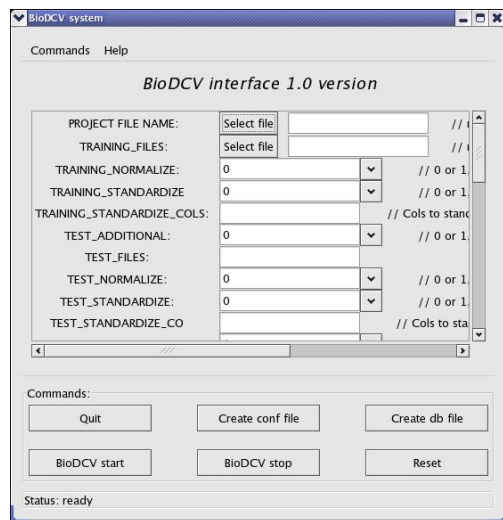
1. Application for analysis of microarray and proteomic data with Support Vector Machine (SVM) classifiers
2. A grid application in production since Spring 2005
3. SQLite database flat files used to manage data/results on grid elements
4. BioDCV is written in C language without external libraries or services
5. Possibility to use LCG2 MPI grid sites
6. **193 CPU days of scalability and footprint tests on microarray data on the INFN – Egrid VO grid infrastructure**
7. **Maldi-TOF proteomics experiments on about 150 grid sites, 220 jobs, 2% failure, on Egee Biomed VO**

Standard LCG user interface commands

- 1.grid-url-copy/lcg-cp db
(from local to SE)
- 2.edg-job-submit BioDCV.jdl (jdl file)
- 3.grid-url-copy/lcg-cp db
(from SE to local)

The BioDCV system





- A Graphics user interface for MS Windows and Linux system (Aug06)
- Next steps for GUI:
 1. create JDL file
 2. link to an LCG2/gLite user interface through a webservice

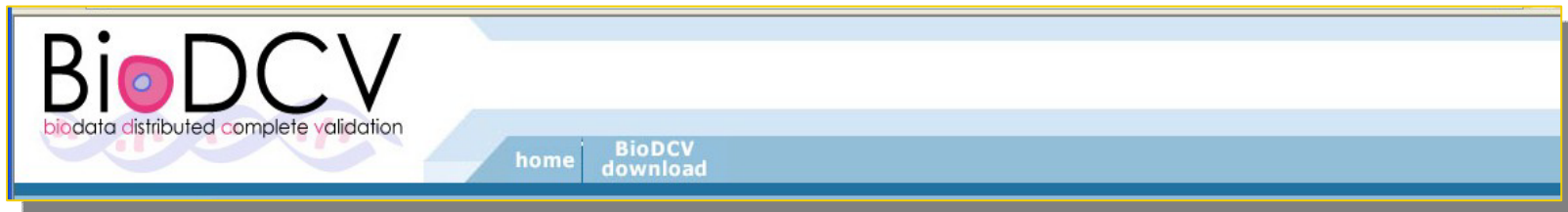
Webservice: How?

1. Apache, mod_python and Zolera SOAP Infrastructure (ZSI)

Why?

1. Use new gLite middleware services
2. Connect BioDCV to other available bioinformatic applications/tools

(first application in: Cannataro M, Barla A, Gallo A, Paoli S, Jurman G, Merler S, Veltri P, Furlanello C. *Workflows, ontologies and standards for unbiased prediction in high-throughput proteomics*, MGED9, Seattle, Sept 2006)



- **SUBVersion repository:** <http://biodcv.itc.it>
- **A production application for predictive profiling on high-throughput technologies**
 - Computational procedures for complete validation
 - Control Selection Bias and Overfitting
 - Stability analysis of Biomarker Lists
- **Applications in molecular oncology**
 - Microarrays: predictive classification
 - Mass spectrometry (SELDI-TOF, MALDI-TOF)
- **Contact:** Cesare Furlanello (furlan@itc.it)