



Enabling Grids for E-science

Introduction to Grid Application Development

Christos Filippidis

(filippidis@inp.demokritos.gr)

Application Support Team

NCSR "Demokritos", Institute of Nuclear Physics

www.eu-egee.org



Information Society



- **Portion of slides (derived from those) prepared by:**
 - Mike Mineter, NESC
 - Charles Loomis, LAL-Orsay
 - Roberto Barbera and his GILDA team
University of Catania and INFN
 - EGEE-II NA4 Activity Member's

- Enabling a whole-system approach
- Collaborative research / engineering / public service ...

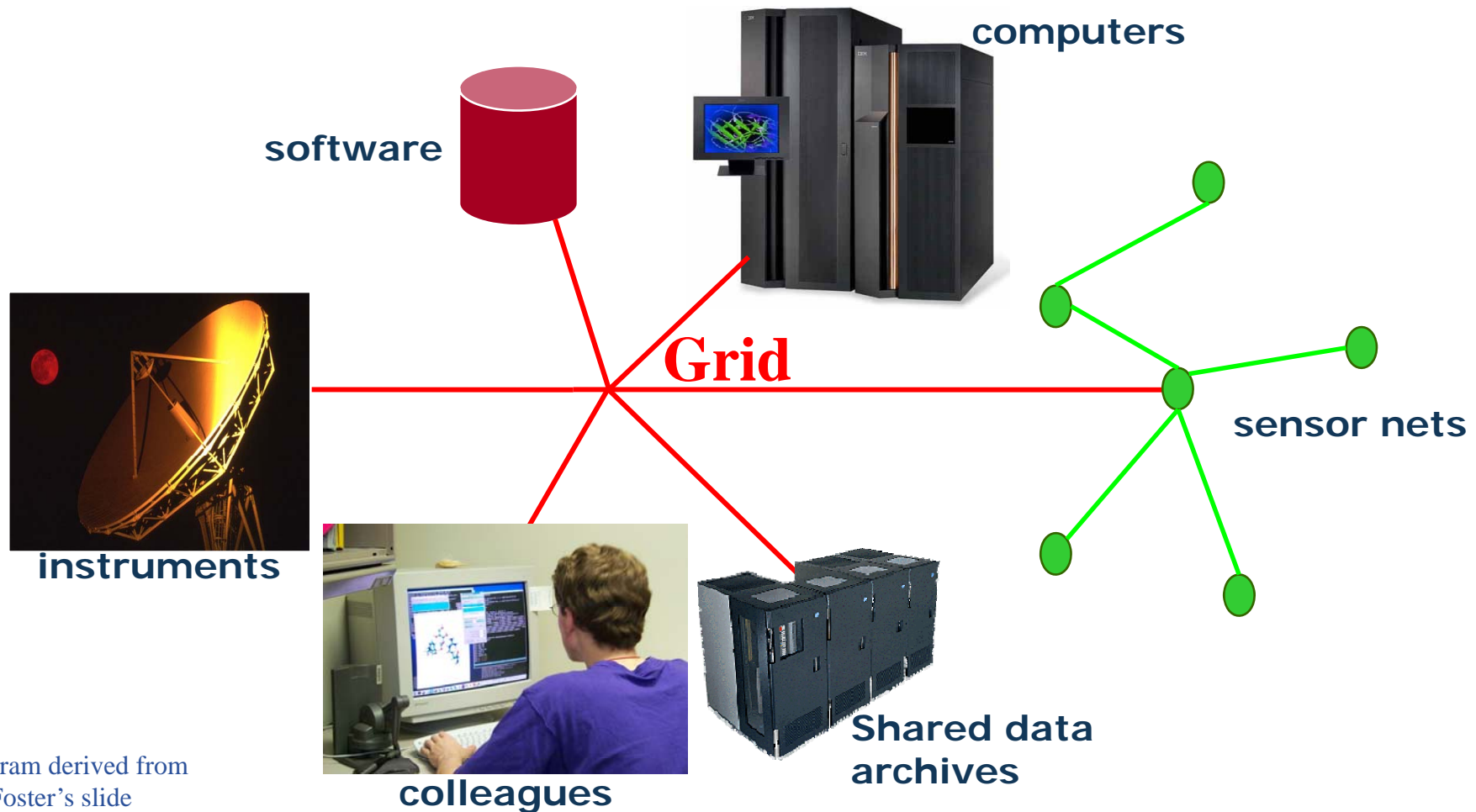


Diagram derived from
Ian Foster's slide

- **I need resources for my research**
 - I need richer functionality
 - MPI, parametric sweeps,...
 - Data and compute services together...

- **I provide an application for (y)our research**
 - How!?
 - Pre-install executables ?
 - Hosting environment?
 - Share data
 - Use it via portal?

- **We provide applications for (y)our research**
 - Also need:
 - Coordination of development
 - Standards
 - ...



Engineering challenges increasing

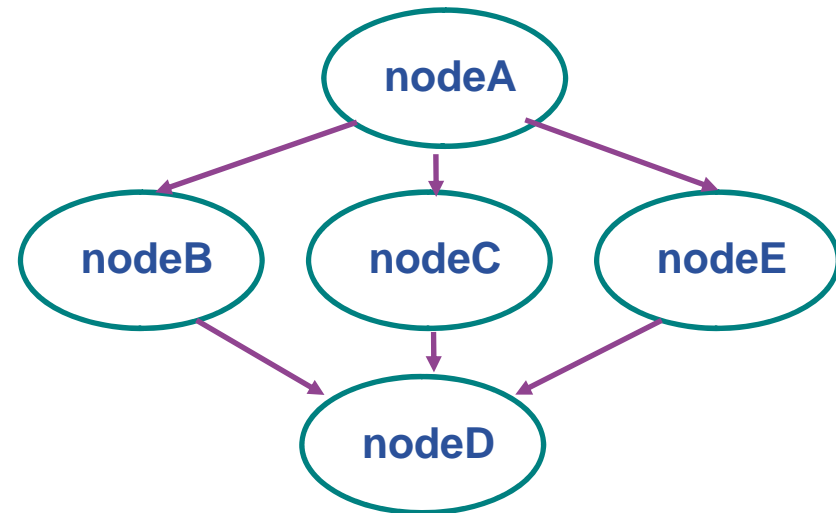
- **Team work!**
- **Engaged in world-wide initiatives – reuse, don't make your own! Cross disciplines for solutions.**
- **From research to production software: ~5 times the effort.**
 - “80% of the time for last 10% of the functionality & reliability”
- **Standardisation is key**
 - For re-use, for dynamic configuration of services,..
 - Both for middleware and domain specific (e.g. GEON)
- **Need to follow a deliberate development process**
 - Waterfall? Rapid prototyping?
 - Requirements engineering, design, implementation, validation, deployment
 - Engaged with the user community

- Application definition
- Environment configuration
- Information discovery
- Data management
- Program definition
- Program execution
- Summary

- Application definition
- Environment configuration
- Information discovery
- Data management
- Program definition
- Program execution
- Summary

1. **Simple jobs – submitted to WMS to run in batch mode**
2. **Job invokes grid services**
 - To read & write files on SE
 - Monitoring
 - For outbound connectivity (interactive jobs)
 - To manage metadata
 - ...
3. **Complex jobs**
 - An environment controls multiple jobs on users' behalf
 - ✂ High-level services
 - ✂ Portals with workflow
 - ✂ Software written for the VO (or by the user)
 - ✂ ...

- **Direct Acyclic Graph (DAG)** is a set of jobs where the input, output, or execution of one or more jobs depends on one or more other jobs
- **A Collection** is a group of jobs with no dependencies
 - basically a collection of JDL's
- **A Parametric job** is a job having one or more attributes in the JDL that vary their values according to parameters
- **Using compound jobs** it is possible to have one shot submission of a (possibly very large, up to thousands) group of jobs
 - Submission time reduction
 - Single call to WMPProxy server
 - Single Authentication and Authorization process
 - Sharing of files between jobs
 - Availability of both a single Job Id to manage the group as a whole and an Id for each single job in the group





Enabling Grids for E-science

Command Line Grid Programming

Spiros Spirou and Vangelis Floros
Greek Application Support Team
NCSR “Demokritos”

www.eu-egee.org

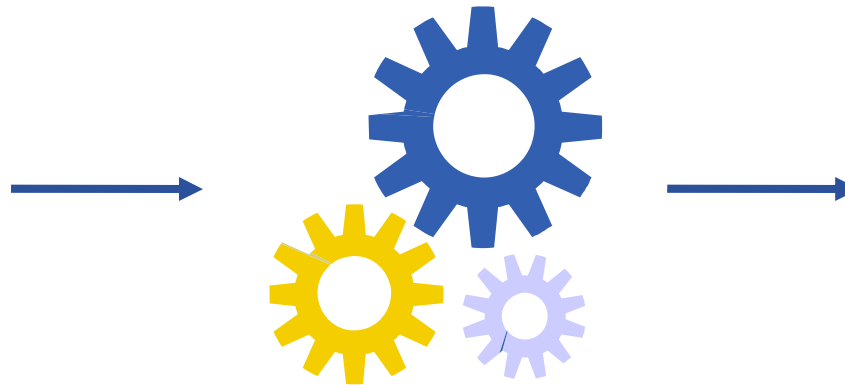


- Application definition
- Environment configuration
- Information discovery
- Data management
- Program definition
- Program execution
- Summary

Image Compression



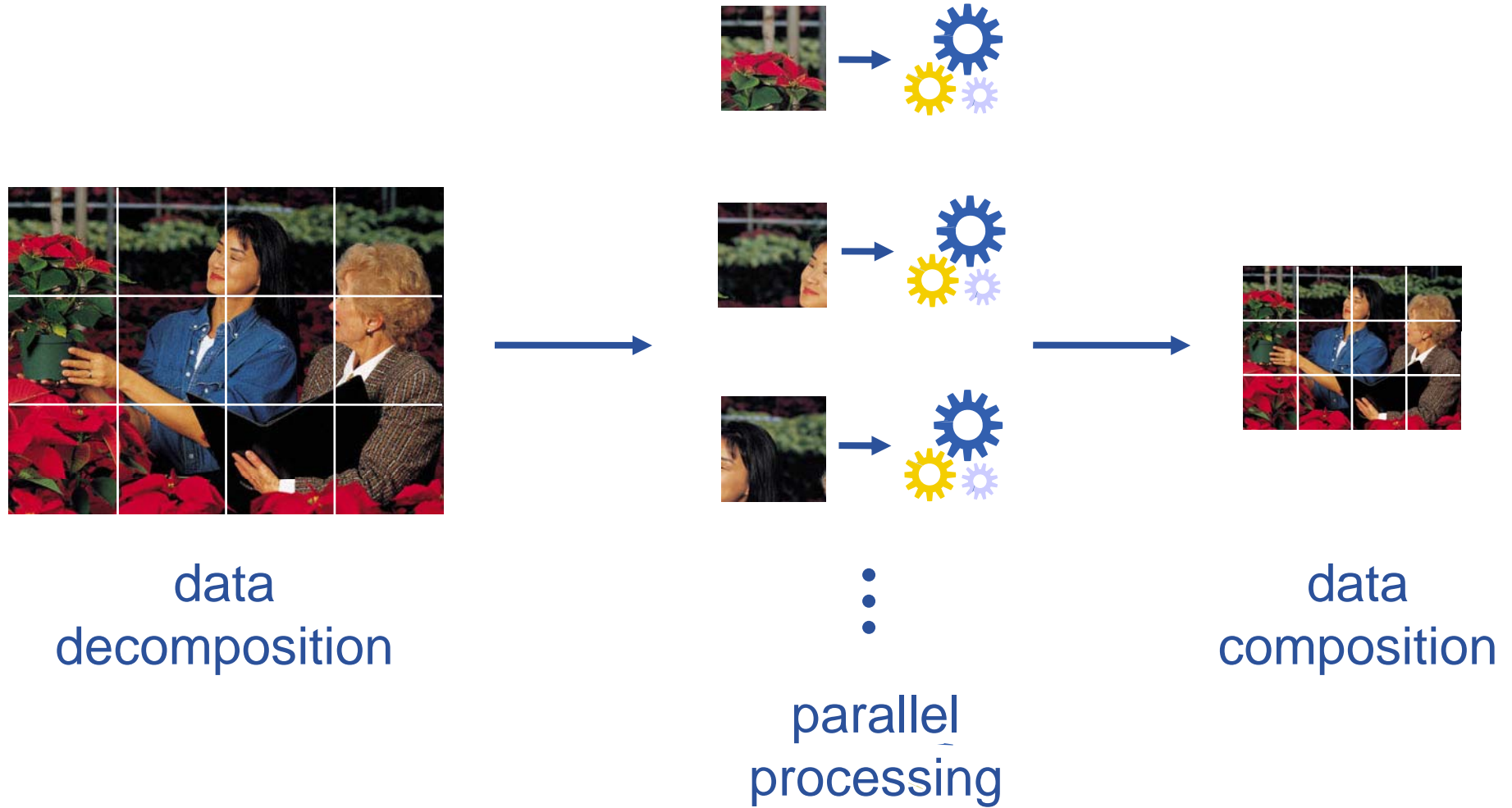
original
image



compression
algorithm

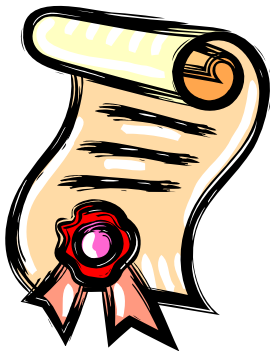


compressed
image



You must already have...

1



CA certificate

2



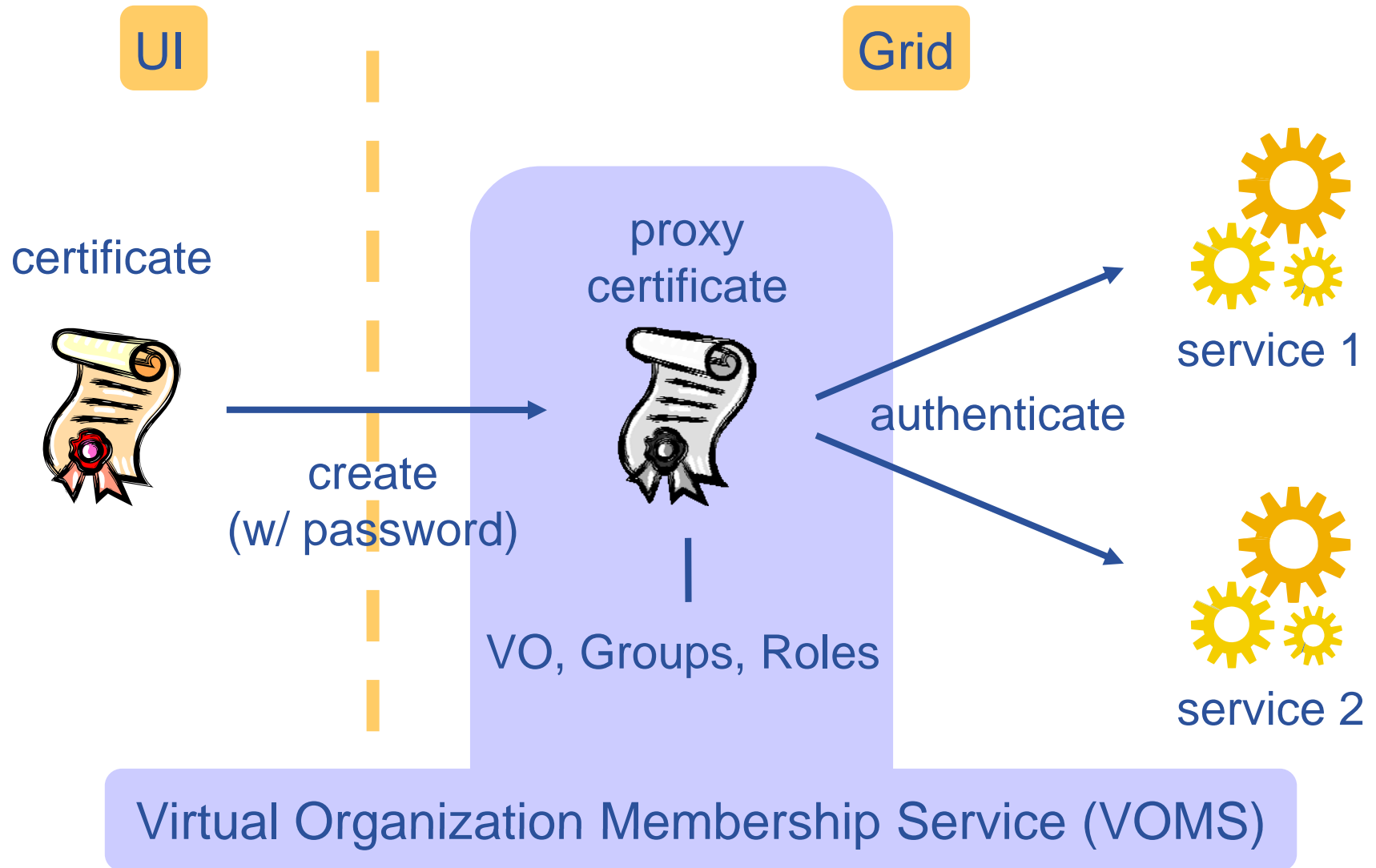
VO membership

3



UI account

- Application definition
- Environment configuration
- Information discovery
- Data management
- Program definition
- Program execution
- Summary



Command (voms-proxy-)	Description
init	Creates VOMS proxy
info	Prints VOMS proxy information
destroy	Destroys VOMS proxy
list	Lists VOMS server attributes

- gLite 3.0 User's Guide
- VOMS Core Services User's Guide

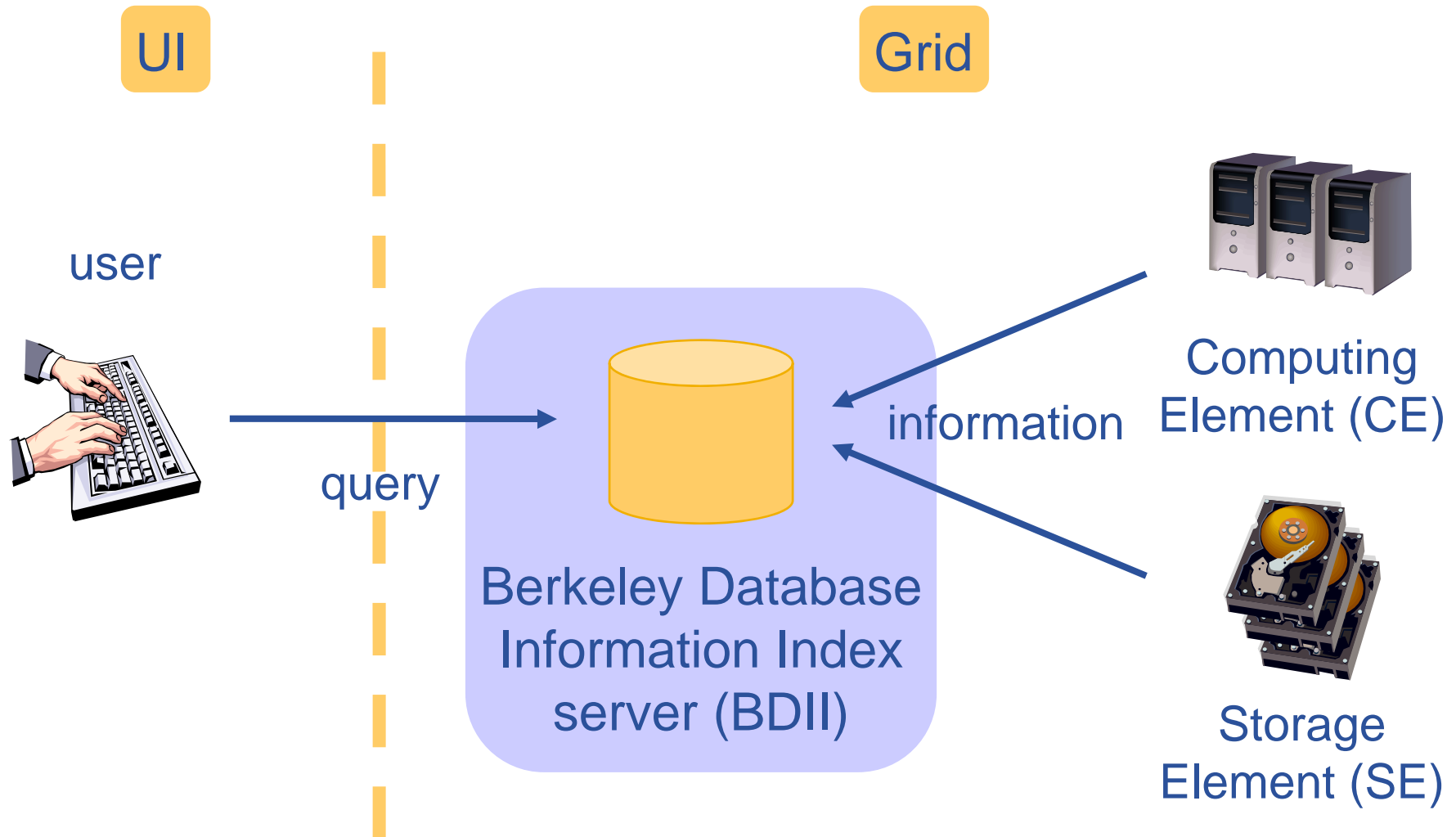
Contact VOMS server and create proxy certificate

```
> voms-proxy-init -voms=hgdemo
```

Check proxy certificate

```
> voms-proxy-info -all
```

- Application definition
- Environment configuration
- Information discovery
- Data management
- Program definition
- Program execution
- Summary



Command	Description
lcg-infosites	Prints Grid service information
Attributes	
--vo	VO name (mandatory)
--is	BDII server name

Options	Description
ce	Prints CE names
se	Prints SE names
closeSE	Prints CE and “close” SE names
all	Prints CE and SE names
lfc	Prints File Catalog (LFC) name

- gLite 3.0 User’s Guide
- lcg-infosites -help

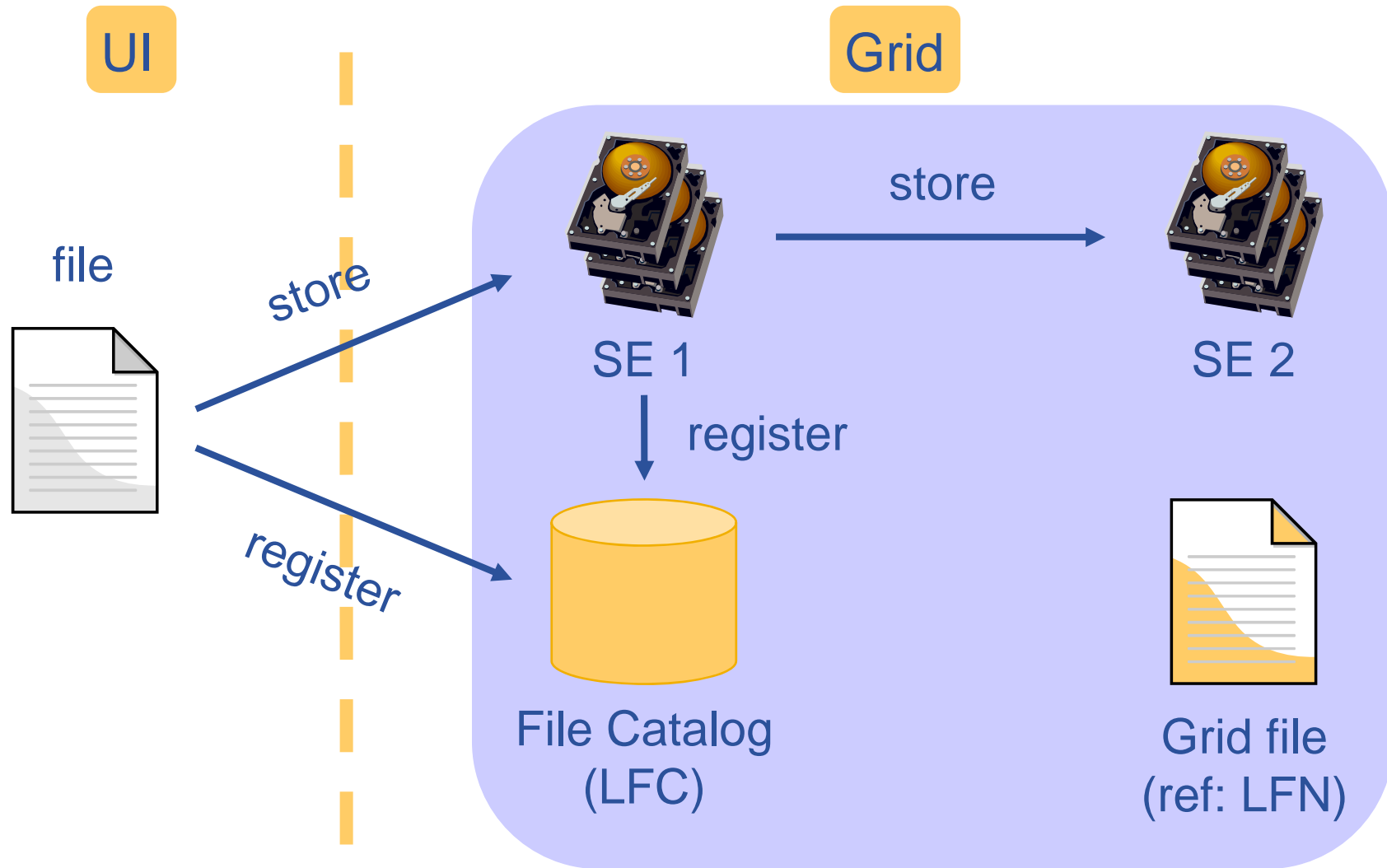
Get CE and SE names for VO

```
> lcg-infosites --vo hgdemo all
```

Get LFC name for VO

```
> lcg-infosites --vo hgdemo lfc
```

- Application definition
- Environment configuration
- Information discovery
- Data management
- Program definition
- Program execution
- Summary



Command (lcg-)	Description
cp	Copies file from Grid to UI
cr	Copies file from UI to Grid
rep	Creates replica of file
lr	Lists replicas of file
del	Deletes file (and replicas)

- gLite 3.0 User's Guide
- lcg-utils man pages

Upload data (sub-images) to Grid

```
> lcg-cr --vo hgdemo -l lfn:/grid/hgdemo/subimg1
  file:/users/johndoe/subimg1
> lcg-cr --vo hgdemo -l lfn:/grid/hgdemo/subimg2
  file:/users/johndoe/subimg2
[...]
```

Locate data (sub-image1)

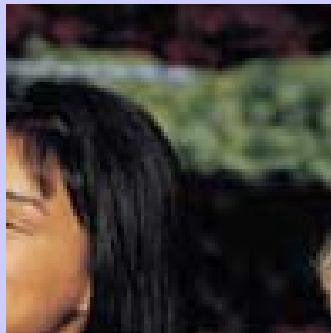
```
> lcg-lr --vo hgdemo lfn:/grid/hgdemo/subimg1
```

Replicate data (sub-image1) to specific SE

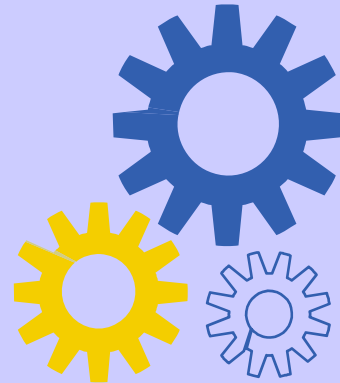
```
> lcg-rep --vo hgdemo -d se01.ariagni.hellasgrid.gr
  lfn: /grid/hgdemo/subimg1
```

- Application definition
- Environment configuration
- Information discovery
- Data management
- Program definition
- Program execution
- Summary

Job Description Language (JDL)



original
input data
sub-image



compression
program
algorithm



compressed
output data
sub-image

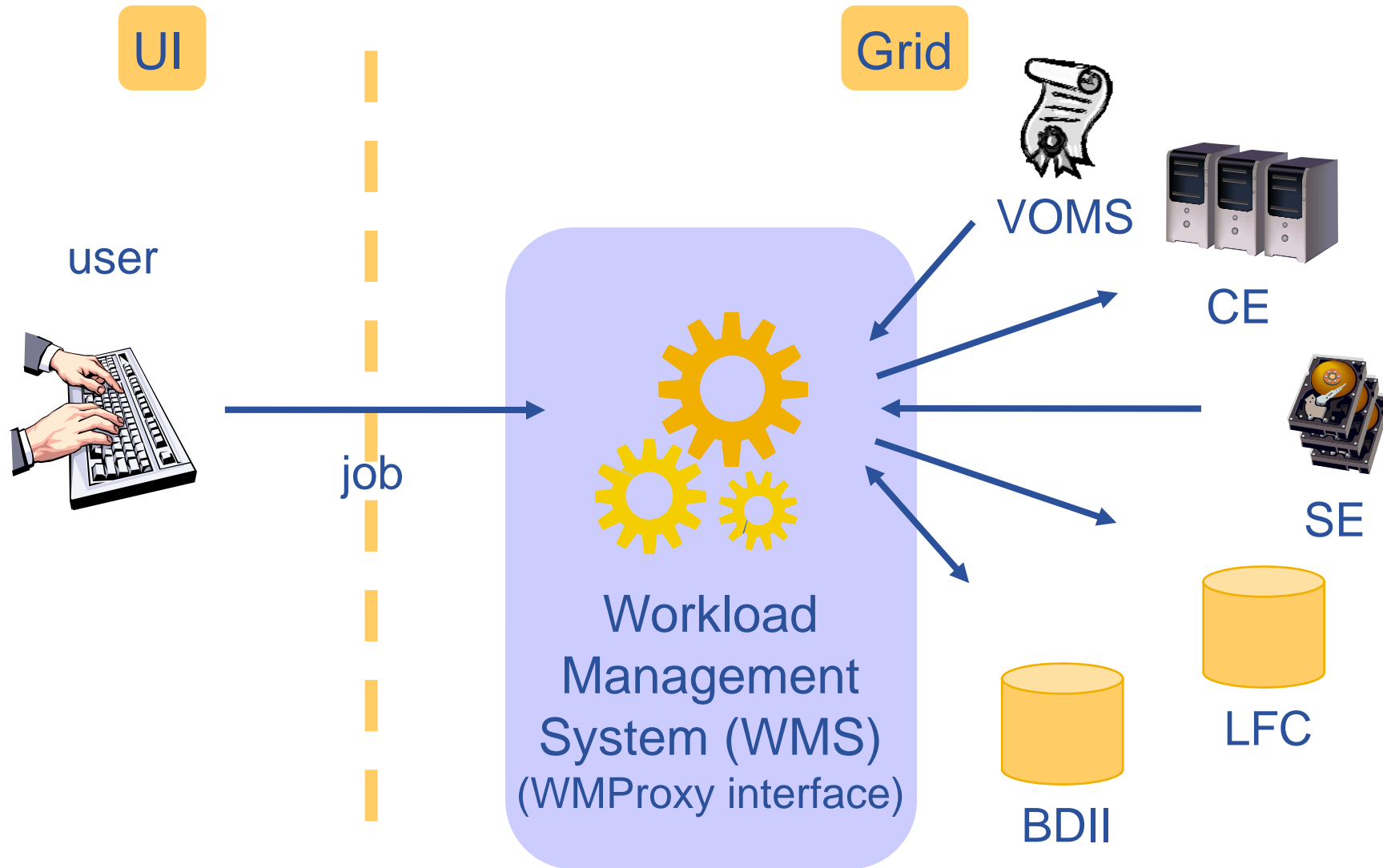
Attribute	Description
Executable	Program to run on Grid
Arguments	Arguments of program
InputSandbox	Files to copy from UI to Grid
StdOutput	File to save program stdout
StdError	File to save program stderr
OutputSandbox	Files to copy from Grid to UI

Attribute	Description
DataRequirements	
InputData	Grid files needed by program
DataCatalogType	Type of catalog referencing Grid files
DataAccessProtocol	Protocol for accessing Grid files
Requirements	Program resource requirements

- gLite 3.0 User's Guide
- JDL Attributes Specification

```
[
  Type = "job";
  JobType = "normal";
  RetryCount = 0;
  ShallowRetryCount = 3;
  Executable = "compress_hgdemo.sh";
  Arguments = "lfn:/grid/hgdemo/egee01/subimg1
  lfn:/grid/hgdemo/egee01/subimg1jpg";
  InputSandbox =
  {"file:///home/training/egee01/tutorial/compress_hgdemo.sh"};
  StdOutput = "std.out";
  StdError = "std.err";
  OutputSandbox = {"std.out", "std.err"};
  DataRequirements = {
    [InputData = {"lfn:/grid/hgdemo/egee01/subimg1",
    "lfn:/grid/hgdemo/egee01/compressjpeg"};
    DataCatalogType = "DLI"];
  };
  DataAccessProtocol = {"gsiftp", "https"};
]
```


- Application definition
- Environment configuration
- Information discovery
- Data management
- Program definition
- Program execution
- Summary



Command (glite-wms-job-)	Description
delegate-proxy	Delegates VOMS proxy to WMPProxy
submit	Submits a job
cancel	Cancels a submitted job
output	Copies OutputSandbox to UI
glite-job-status	Prints submitted job status

- gLite 3.0 User's Guide
- WMPProxy Service User's Guide

Submit job and save job id (automatic delegation)

```
> glite-wms-job-submit -a -o job_id  
compress_simple.jdl
```

Check job status

```
> glite-job-status -i job_id
```

Get job output when done

```
> glite-wms-job-output -i job_id
```

- Application definition
- Environment configuration
- Information discovery
- Data management
- Program definition
- Program execution
- Summary

- Parallelize application
- Authenticate with VOMS
- Get information with lcg-infosites
- Upload data with lcg-utils
- Describe application with JDL
- Execute application with glite-wms-job-*

The Grid is fast, simple, and free. So, use it!



Enabling Grids for E-science

Questions?

Christos Filippidis

(filippidis@inp.demokritos.gr)

Application Support Team

NCSR "Demokritos", Institute of Nuclear Physics

www.eu-egee.org



Information Society





Enabling Grids for E-scienceE

http://wiki.egee-see.org/index.php/Programming_from_the_Command_Line

www.eu-egee.org

