Clinical Decision Support Systems (CDSS) in GRID Environments

HealthGrid Conference
Oxford, April 6-8, 2005

Ignacio Blanquer
Universidad Politécnica de Valencia
Departamento de Sistemas Informáticos y Computación (DSIC), Valencia, Spain.

www.eu-egee.org
Contents

• What is a CDSS?
• User Community
• Objectives of the Grid Approach
• Components of the Grid Application
• Applications
  – Searching an Available Classification Engine
  – Executing the Classification Process
• Future Work
• Conclusions
What is a CDSS?

• Application that Extracts Medically Relevant Knowledge from a Large Set of Information with the Objective of Guiding the Practitioners in their Clinical Practice.

• It Consists on Several Trained Classification Engines that Process a Formatted Input Determining its Category from a Predefined Set.

• Currently 7 Classification Engines and Three Application Areas (Talassemia, Soft Tissue Tumours and Schizophrenia).
• **Usage of the Tool**
  – For the Automatic Classification of a Large Set of Cases (Epidemiology Studies), Assuming the Classification Error.
  – For an Individual Case when More Information is Needed, Under the Supervision of the User.

• **Medical Users: Depending on the Area**
  – Anaemia
    ▪ Haematological Department of Hospital Dr. Peset.
  – Soft Tissue Tumours Classification
    ▪ eTumour Consortium (http://www.etumour.net/).
    ▪ ADIRM (Asociación Española para el Desarrollo y la Investigación en Resonancia Magnética).
Objectives of the Grid Approach

- **Problems and Needs**
  - Remote Usage. Sending the Classification SW Will Compromise IPR.
  - Security and Accounting.
  - Batch Usage in Epidemiology Studies.
  - Publication of Resources and Features to Select the Most Suitable One.

- **Advantages of Grid**
  - More Natural Concept for Services.
  - Security and Accounting by Definition.
  - Access to Higher Resources in Batch Usage: Advantages on the Combination of Results from Different Sources.
  - Robust and Dynamic Publication of Available Resources.
• **Components**
  – Gate-to-Grid.
  – Set of Trained Classifiers (Engines).
  – A Searching System.
  – Execution of Engines with a User-Specific Input.
  – Security.
Grid Architecture (Gate to Grid)

- Need for a Gateway to the Grid (Gate to Grid).
- The Gate to Grid is the Server that Provides the Users of the CDSS Applications with a User-Friendly and Secure Web Vision for the Grid Environment.
- It has two Interfaces:
  - Web Interface (Web Services) \(\rightarrow\) Interacts with the End Users.
  - User Interface of LCG (UI) \(\rightarrow\) Interacts with the Grid.
- It Translates Operations from Web Services to LCG Grid Commands Through a Secure Channel.
The Different Processing Nodes Provide a Set of Trained Classifiers (Engines).
- Each Engine has a Descriptive Report of all its Features and Relevant Information in an XML Document.

The Processing Nodes are the Computers Elements in the LCG Environment.

Relevant Information from the Engines are Published with the Monitoring and Discovery System (MDS) of LCG → It Gets Automatically the Information from the Descriptive Reports of Each Engine.

Different Engines can be Installed in the Same CE.
Grid Architecture (Searching & Execution System)

- **Searching System**
  - The End User Interacts Through Web Services.
  - A Searching System Enables to Locate the CEs Containing Engines According to a Criterion (Corpus, Efficacy, etc…).
  - The Different Criteria are Published Thought the MDS and Chooses the Information of the XML Reports.

- **Execution System**
  - The End User Interacts Through Web Services.
  - The Execution of Engines with a User-Specific Input.
    - First → Choose the Classifier Engine with the Searching System.
    - Second → Execute the Classifier Engine Selected in the GRID with the Given Parameters.
  - Get the Results in the form of an XML Document.
Grid Architecture (Searching)

- **Searching Defines:**
  - The Corpus.
  - The Classification Criteria.
  - The Efficacy.
- **Multiple Conditions can Be Specified and Linked Through AND / OR Operations.**
- **Only the Engines and Sites that Fulfil the Criteria are Shown on the Left Panel.**
• Depending on the Corpus, Different Input Information is Presented.

• Different Engines for the Same Corpus Share the Same Interface, Although Performance Could be Different.

• Multiple Engines Can be Used Simultaneously Providing a Consolidated Result.
**Grid Architecture (Execution)**

- **Example: Node Selection and Filling of Parameters.**
- **Submit Button**
  - Generation of the XML Request.
  - Translation of the XML Request in a JDL File.
    - Executable="launcher.sh";
    - Arguments="Mentzer arg_file temporal_output.job temporal_err.err";
    - stdOutput="std.out";
    - stdError="std.err";
    - InputSandbox={"arg_file"}
  - Submission of the Job.
    - edg-job-submit -h <host> jdl_file -o job_id
  - Retrieval of Results.
    - edg-get-results -i job_id
  - Translation of Results to an XML File.
Two Parts Must be Distinguished

- Security of Web Services
  - Identification of the User and Reliability of the Authenticity of the Services.

- Security of Grid Environment
  - Authorization of the Access to the System Through the VO.
  - Mapping of the Users.
Grid Architecture (Security)

- **Security of Web Services**
  - The Protocol Used by the Web Services is HTTPS, which is Based in SSL (Secure Sockets Layers).
  - The End User Provides a User Certificate (X.509 CA DataGRID).
  - The Gate-to-Grid (Web Service Container) is Certificated with a Server Certificate (X.509).

- **Security of Grid Environment**
  - Based in GSI (Grid Security Infrastructure).
  - GSI is Based in X.509 Certificates and SSL Protocol.
  - The DN of the User Able to Reach the Web Service is Obtained and Compared with the VO Mapfile.
  - If Accepted, the Grid User Creates a Proxy for Launching the LCG Grid Commands.
  - All Valid End-users (CA DataGRID) are Mapped to a Single Grid User for Launching LCG Grid Commands.
Grid Architecture (Security)

- The Two parts are Connected Through a User-Mapping Process in the Gate-to-Grid.
- All Web-users are Mapped to a Single Grid-User.

![Diagram of Grid Architecture](image)
Further Work

• Integration of the SW Library on the Tool used in eTumour Consortium (in Process).
  – Enabling Batch Processing.
  – Enabling Multi-Classifier Classification.

• Integrate the Training Part on the SW Library (in Process).

• Provide the Searching of Properties Through R-GMA Rather than MDS.

• Extend the System to Other Cases.
  – The Public Health Unit of the Valencia Region is Interested.
    ▪ Clinical Records are of the order of 3 Million per Day.
    ▪ Depending on the Disease to Test (Initially Arthritis), Relevant Records are of Several Thousands per day..
  – Relation of Schizophrenia and Genomic Factors.
Conclusions

• The System is Installed and Running on LCG2.
• The Development of New Classification Components in a Configured Site is Easy.
• Individual Jobs are Short, but Classification of Large Number of Jobs is Relevant in Epidemiology Studies.
• Medical Interest is on-Line with the Priorities of Research in eHealth of the EU.