

# Building Grid Infrastructure for Biomedical Applications

## The MediGrid Project

Ondřej Krajíček, krajicek@ics.muni.cz

**Institute of Computer Science and Faculty of Informatics**  
Masaryk University

CoreGrid

7th July 2006

# The MediGrid Project

- focuses on establishing grids in biomedicine as utility to support biomedical experts (physicians, surgeons, chemists, etc.) – biomedical grids
- demonstrate and explore the possibilities of deploying biomedical grids in everyday use

- what is biomedicine?
- biomedical knowledge and biomedical applications
- do we need biomedical grid?
- leveraging grids

# What is biomedicine?

- not a question for computer scientists or engineers
- biomedicine = biochemistry + biology + medicine
- research discipline with practical applications: focus is on diagnostics

- how IT is used?
  - various biomedical ontologies to capture semantics of terms (ontological databases)
  - isolated information systems used in hospitals and by specialists to track patient records
  - knowledge published in journals and scattered across the Web
  - communication via e-mail, conferencing tools sometimes

# Do we need Grids in/for biomedicine?

- biomedical experts need platform for:
  - secure sharing of data (e.g. patient records) and their anonymisation
  - communication and collaboration
  - resource sharing
  - **knowledge publication and sharing**
- all this with:
  - security and fine-grained access control
  - accounting
  - resolved legal issues or at least no significant design flaws preventing this
  - with all the advantages of grid computing

- **ontologies**
- **indicators**
- **modules**

- capture (label) semantics of subjects
- explicitly defined (by the owner, author, physician, etc.)
- automatically derived (based on data relationships, patterns, rules, etc.)

- **well** and properly **defined data**, often published and **reviewed**
- semantics of indicators is described using ontologies
- not necessarily atomic or elementary data
- not necessarily (directly) measurable
- indicator definitions are captured in *indicator classes*
- examples of indicators: body height, blood pressure, etc.

- “biomedical algorithms”
- formulae, processes, algorithms
- well and properly defined, often published and reviewed
- operations on indicators

# Formalising Definitions

- module and indicator definitions
- human readable part:
  - name, description of the purpose and mechanisms
  - implementation details (even source code)
  - formal proofs (if any necessary)
  - anything appropriate

## Formalising Definitions (2)

- machine processable part:
  - syntactic definition: data types (I/O indicators in case of modules)
  - semantic annotation
  - QoS properties
  - references and reviews
  - authentication information
- everything has secure identity: non repudiation and integrity are of concern

# Typical Use-Cases

- 1 indicator classes and required modules are known, user wants to do the computation “somehow” and “somewhere” (naturally extends to the SPMD scenario)
  - 2 input indicators are known, user wants to discover (all/partially restricted) suitable modules
  - 3 required (output) indicators are known, users wants to discover what indicators must be supplied
- multiple ways to compute particular indicators – selection: trust, QoS, **semantics**
  - syntactic and semantic matching
  - **data-driven usage** – indicators

# Leveraging Grids

- communication – collaborative grids (access grids)
- knowledge publication and sharing – semantic grids, computational and data grids
- user interface – grid portals (e.g. P-GRADE)
- module implementation – service oriented grids, GEMLCA

# Leveraging Grids (2)

- many aspects of biomedical applications
  - 1 discovery – implement distributed repository for modules and indicator classes
  - 2 selection – QoS, syntactic and/or semantic matching
  - 3 computation – workflows (to compose modules where appropriate), data grids to manage data storage, computational grids to manage computations
  - 4 monitoring – need for service-oriented monitoring infrastructure (we are working on it)
  - 5 users – end users, authors, reviewers

# Ideal Scenario – End User

- 1** expert (physician, surgeon, etc.) identifies a task she needs to solve in terms of indicators
- 2** expert uses grid portal to select modules (enact workflows, etc.) – this process is aided by the portal, expert needs only to confirm proposed solutions
- 3** expert supplies input data (identifies the data source) and executes the computation

# Ideal Scenario – Module Publisher

- 1 a researcher develops a new technique of computing or measuring an indicator
- 2 the researcher creates module and indicator definitions
- 3 the researcher builds a team (or does himself) which implements everything
- 4 the implemented module is published in the infrastructure and becomes available to others

# Ideal Scenario – Rewiever

- 1 a reviewer evaluates some module, its definition, implementation, indicator definition, etc.
- 2 the reviewer publishes a review which contains identification of the reviewed subject (using secure identity)
- 3 the module author/publisher may include the reference to this review as part of the module definition

# Conclusions

- 1 biomedicine presents interesting applications for Grid computing
- 2 not computationally intensive but needs other grids as well (data grid, access grids, semantic grids)
- 3 key implementation issues arise from the way how biomedical experts see computers and how computer scientists/engineers see biomedicine
- 4 iterative prototyping