OpenMolGRID
Open Computing GRID for Molecular Science and Engineering

Forschungszentrum Jülich
in der Helmholtz-Gemeinschaft

OpenMolGRID:
Complex Problem Solving in Molecular Design

EUROGRID Workshop
Cracow, October 29, 2003

Mathilde Romberg
Bernd Schuller
Forschungszentrum Jülich

Outline

- The OpenMolGRID Project
- Molecular Design and Engineering
- Grid Architecture
  - Integration of Applications
  - Database Access
  - Workflow Support
- Status and Outlook

OpenMolGRID

- Funded in part by EC: IST-2001-37238
- 01.09.2002- 30.11.2004
- Partners:
  - University of Tartu, EE (Project Coordinator)
  - University of Ulster, UK
  - Mario Negri Institute, IT
  - Forschungszentrum Jülich, DE
  - ComGenex Inc., HU
- www.openmolgrid.org

Project Objectives

- Development of tools for secure and seamless access to distributed information and computational methods relevant to molecular engineering within the UNICORE frame
- Provide a realistic testbed and reference application in life science
- Development of a toxicity prediction model validated with a large experimental set
- Provide design principles for next-generation molecular engineering systems

Work Packages

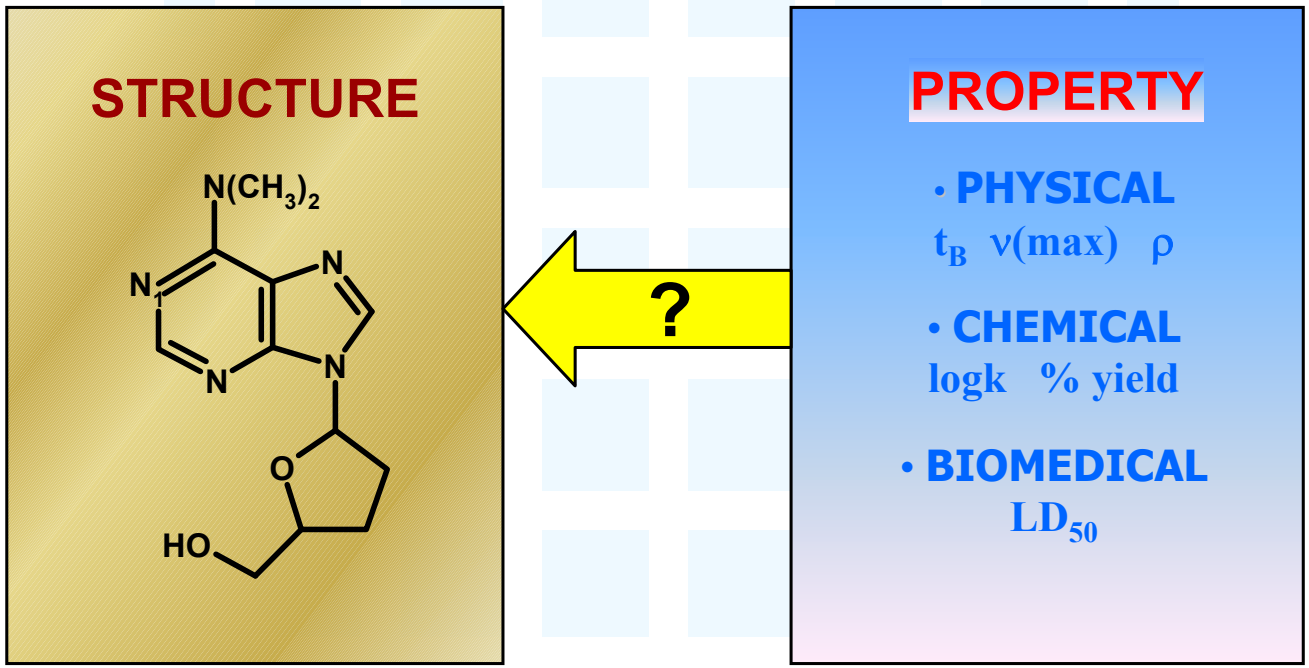
- WP1: Grid Data Warehousing of Molecular Structure -- Property (Activity) Information
- WP2: Molecular Descriptor Generation and QSPR Model Building on the Grid
- WP3: Computational Molecular Engineering of New Compounds and Materials
- WP4: Grid Integration
- WP5: Test application of the OpenMolGRID System for Chemical and Pharmaceutical Predictions

Molecular Engineering

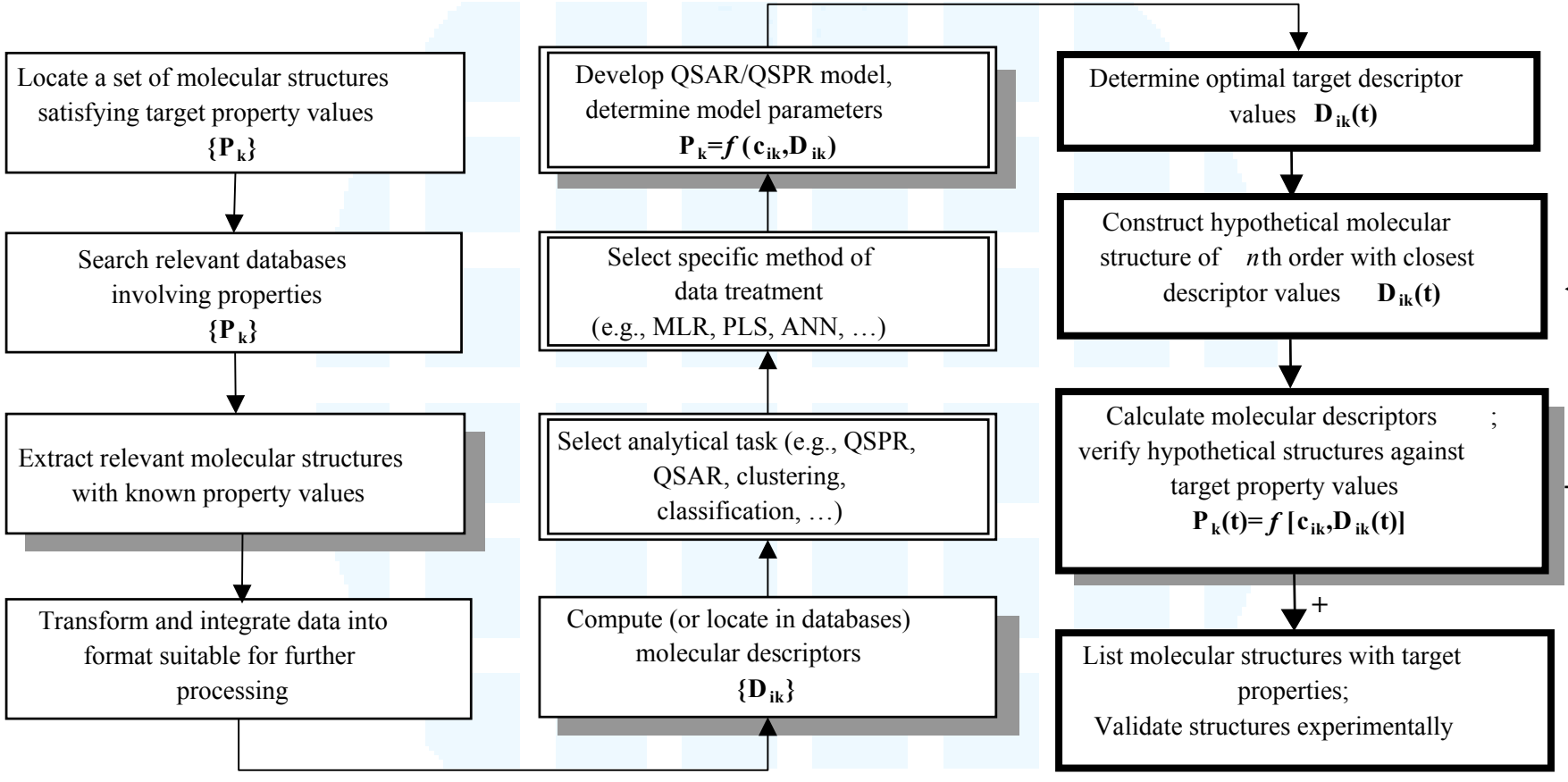
Rational design and targeted synthesis of

- new molecules
- new materials
- new chemical reactions
- new technological processes

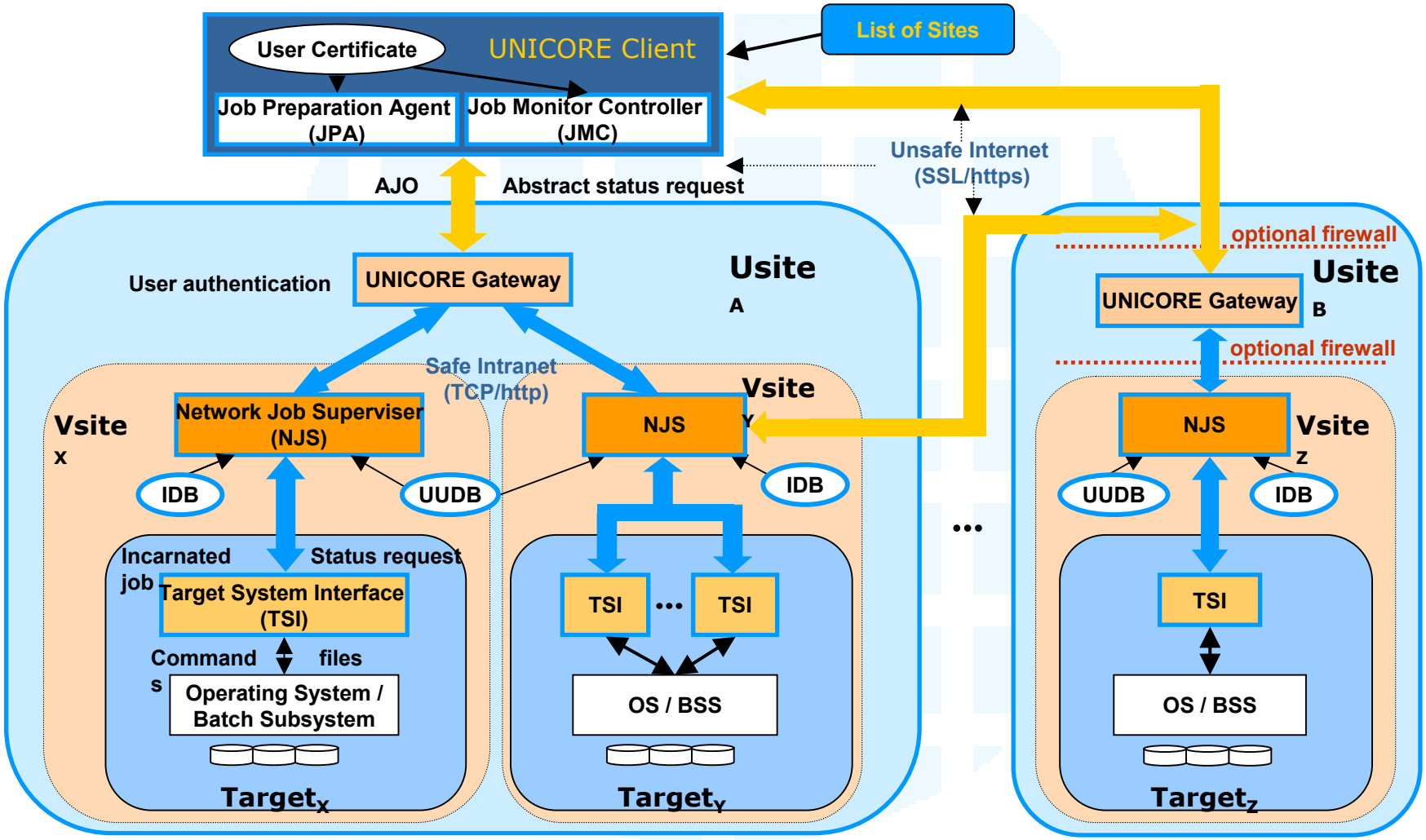
Molecular Engineering



Molecular Design



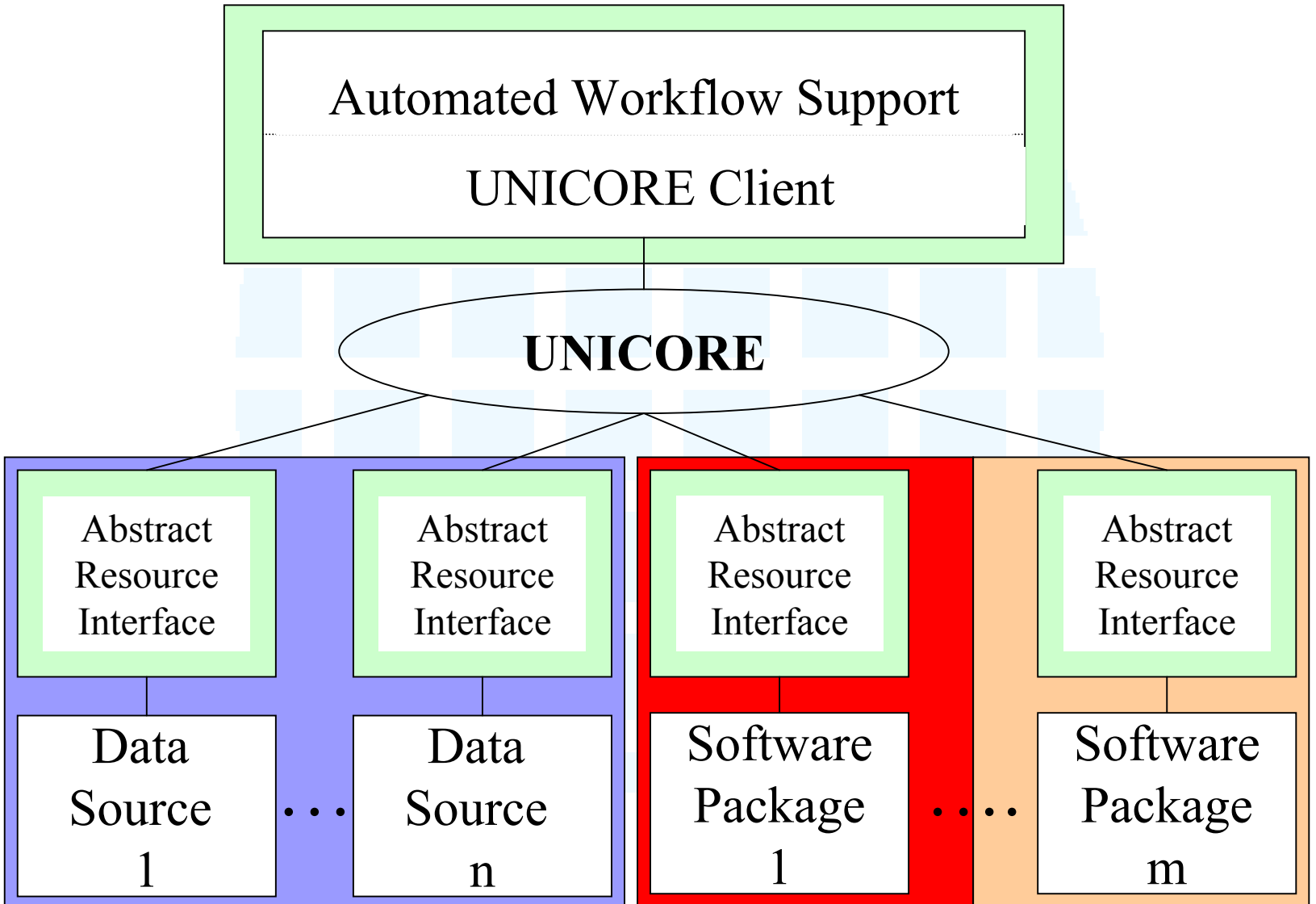
Basis: UNICORE Infrastructure



OpenMolGRID enhances UNICORE

- Plugins for
  - Classes of applications for molecular calculations
  - Workflow support
  - Database access
- Application ‘Database Access Tool’
  - Interface between UNICORE and database
  - Flexible output formats (XML, XSLT)
- Abstraction Layer for software modules

OpenMolGRID Architecture



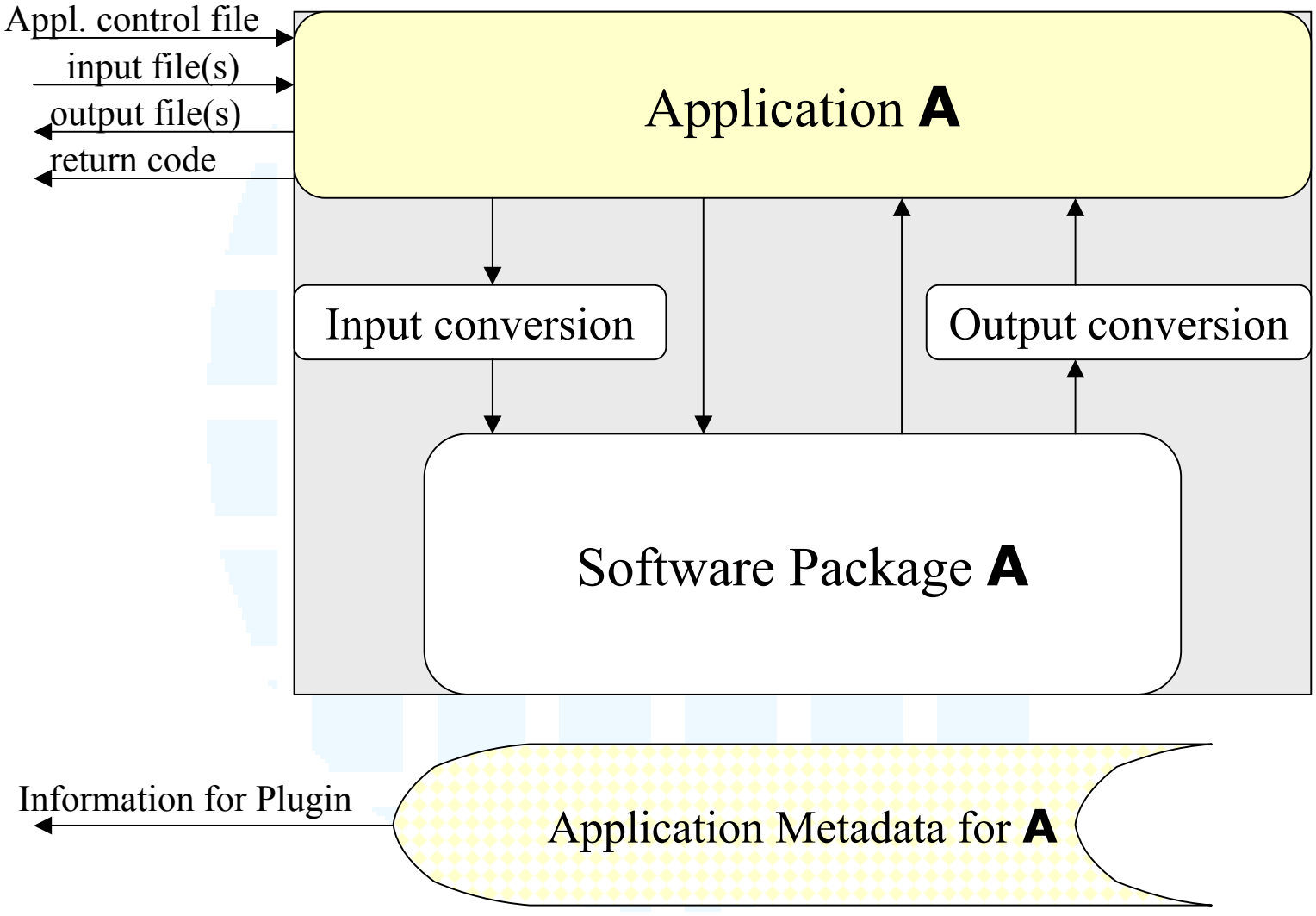
Application Specific Support

- Client plugins
  - GUIs for applications and workflow
  - Resource selection
- Resource definition for applications
  - Part of Incarnation DataBase
- Application metadata
  - Description of the application
  - Information for the client plugin

Application Definition in IDB

- APPLICATION *A n.m metadata_file*
- INVOCATION *A_n.m* [……]
- Metadata format:
  - Task (name, description)
  - Input [infile (type, use)]*
  - Output [outfile (type, occurs)]*
  - Appspecific *information for client plugin*

Application layer

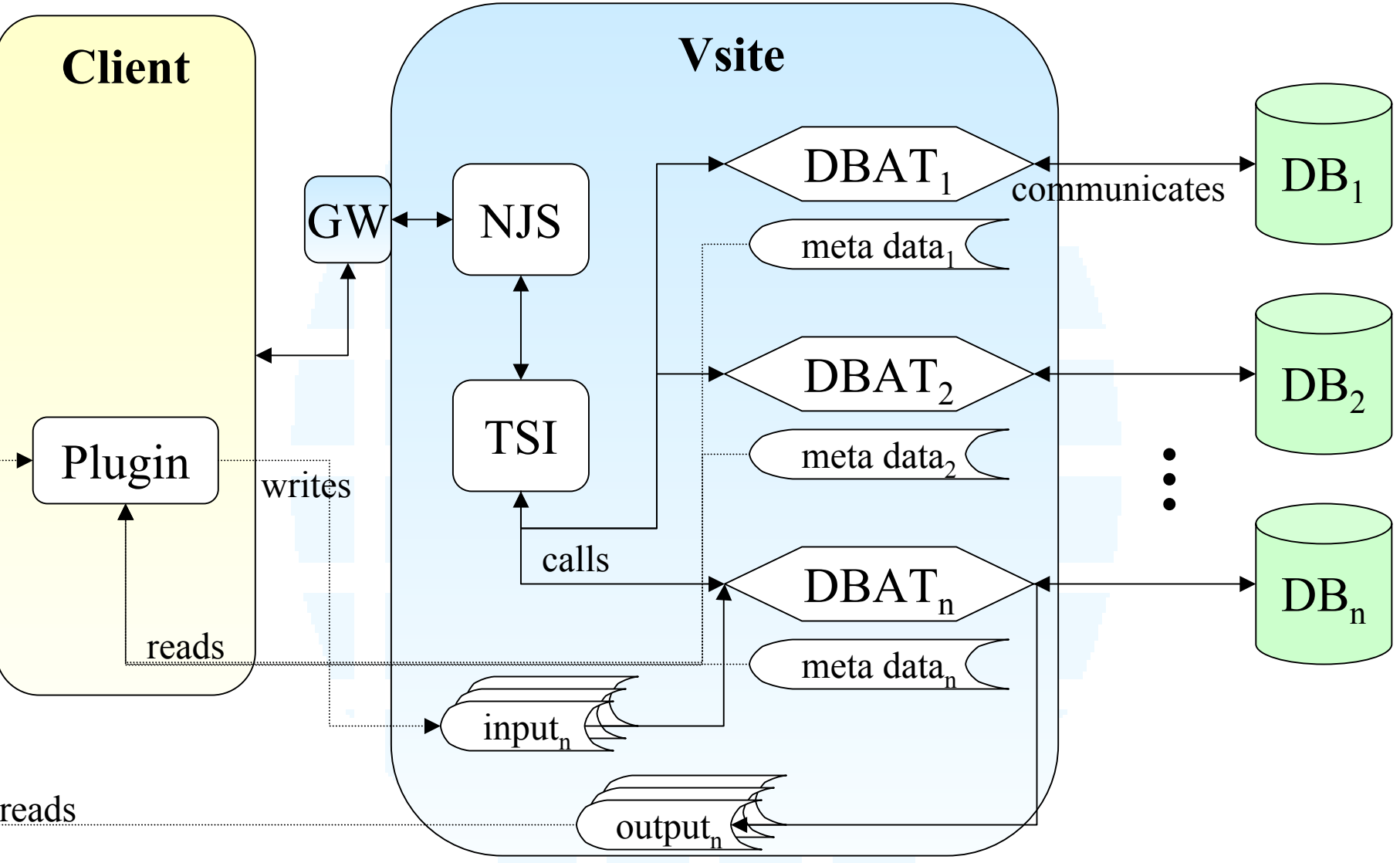


Database Access Architecture

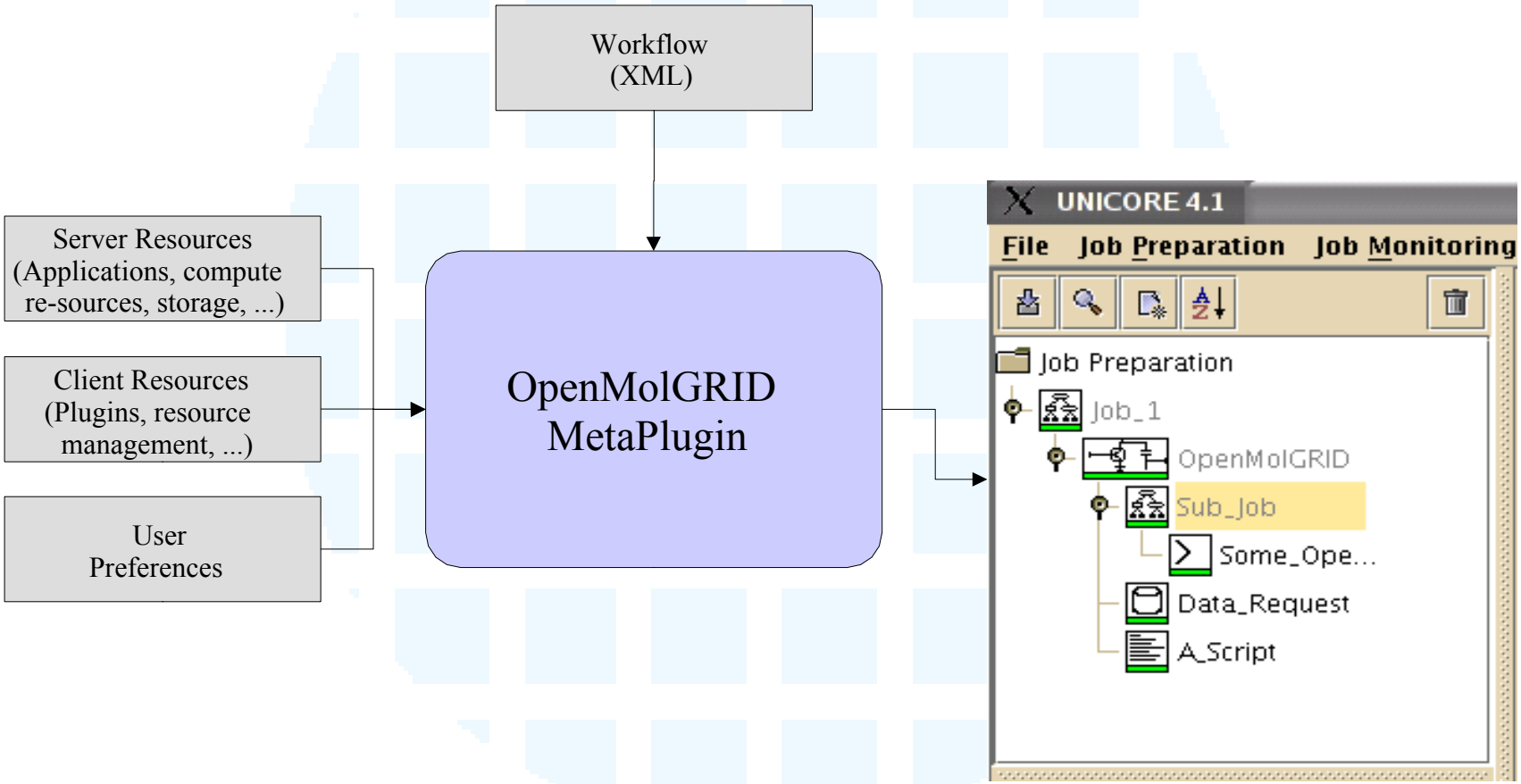
- Access to a data source is seen as an application
- Structural information is described in the metadata file
- User interface is an application specific interface

Database Access Architecture (cont.)

- General
  - Valid for all types of information
  - Valid for all kinds of databases
- Flexible
  - Metadata file allows for adaptation to db changes
- Extensible
  - Arbitrary output formats can be supported
  - XML Document Type Definition


Client
Vsite
GW
NJS
TSI
Plugin
writes
reads
calls
DBAT1
DBAT2
DBATn
communicates
meta data1
meta data2
meta datan
DB1
DB2
DBn
inputn
outputn
reads

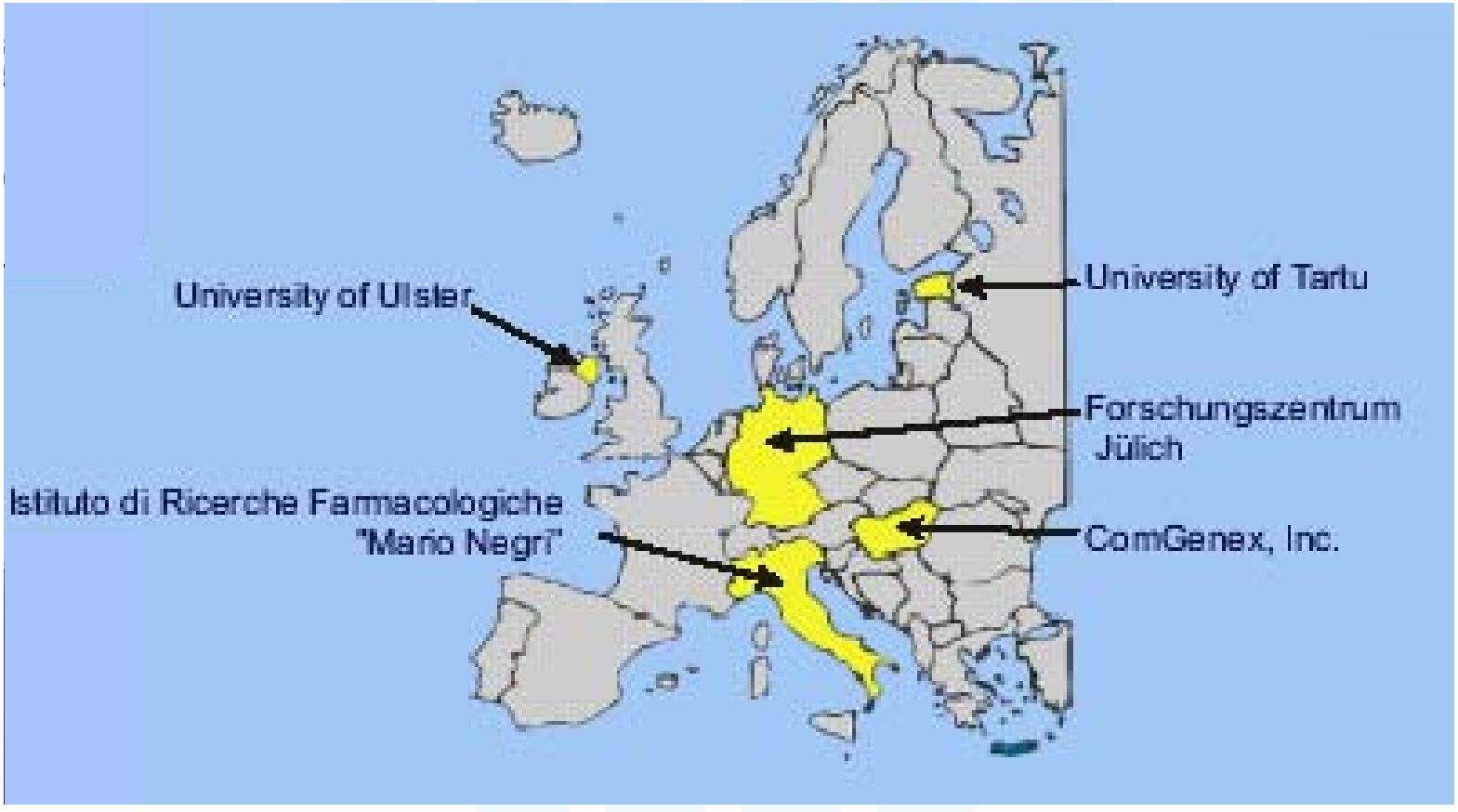
Workflow Support



Meta-Plugin

- Plugin able to see all task plugins
- Workflow description (XML) is used to generate UNICORE job tree
- Look for matches between output file / input file specifications of two dependent applications
- Allow for user intervention at predetermined positions in the workflow (hold forever / release)
- Distribute tasks to multiple Vsites
- Select target site(s) and application resource
- Insert transfer and data conversion tasks where necessary

Testbed


University of Ulster
University of Tartu
Forschungszentrum Jülich
Istituto di Ricerche Farmacologiche "Mario Negri"
ComGenex, Inc.

Status

- Data Warehouse (MOLDW) specified
- Application interfaces in progress
- Abstraction layer (DBAT) for relevant Databases available
- Initial version of Meta-Plugin available
- Initial testbed set up
- Workflow specification for selected steps available

Outlook

- Feed MOLDW with data
- Develop abstract resource interfaces for
  - descriptor calculation
  - model development
- Develop resource information provider plugin