
B2A Pharma Prototype

Implementation of an industrial-strength pharmaceutical workflow in a Grid environment

Falk Zimmermann
NEC Europe Ltd.
IT Research Division – Sankt Augustin, Germany

October, 2nd 2007
EGEE Conference
Budapest

SIMDAT Facts

- EU Information Society Techn
- GRID Project
- Duration: 4 years

- Start date: September 1st
- 28 partners



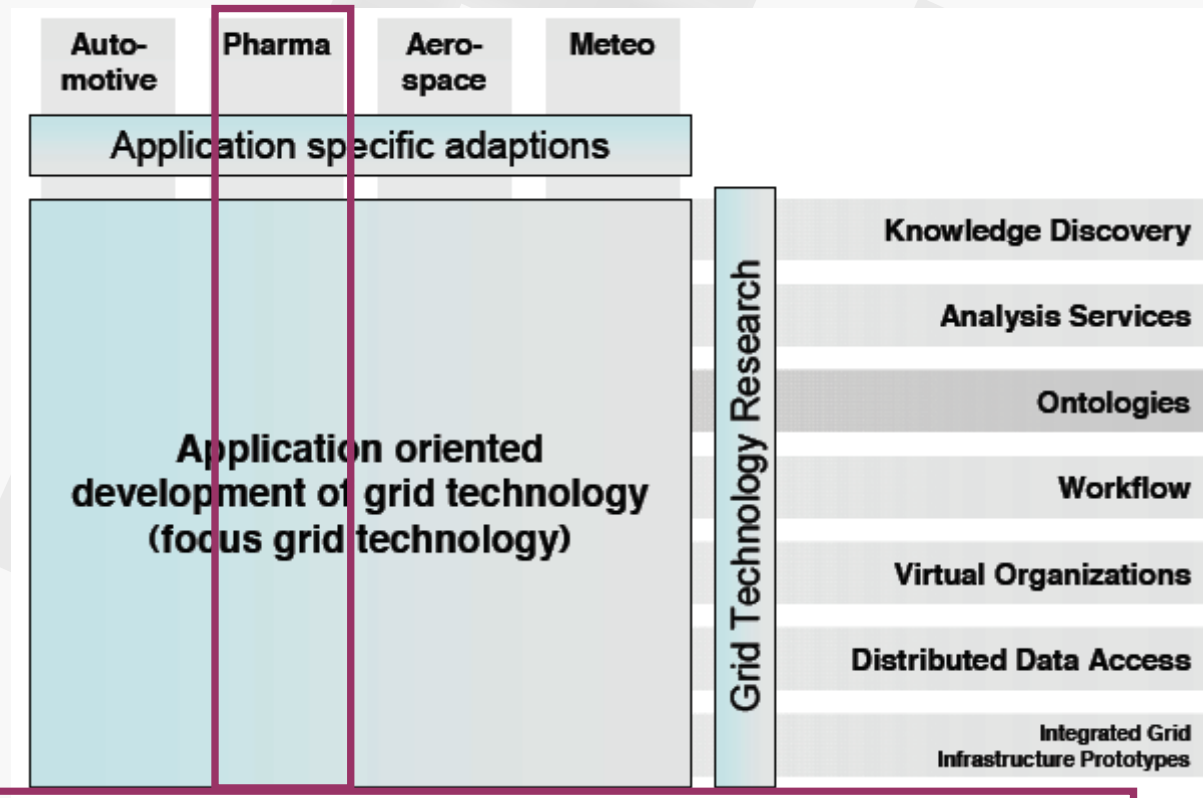
- Start date: September 1st
- 28 partners



SIMDAT Objectives

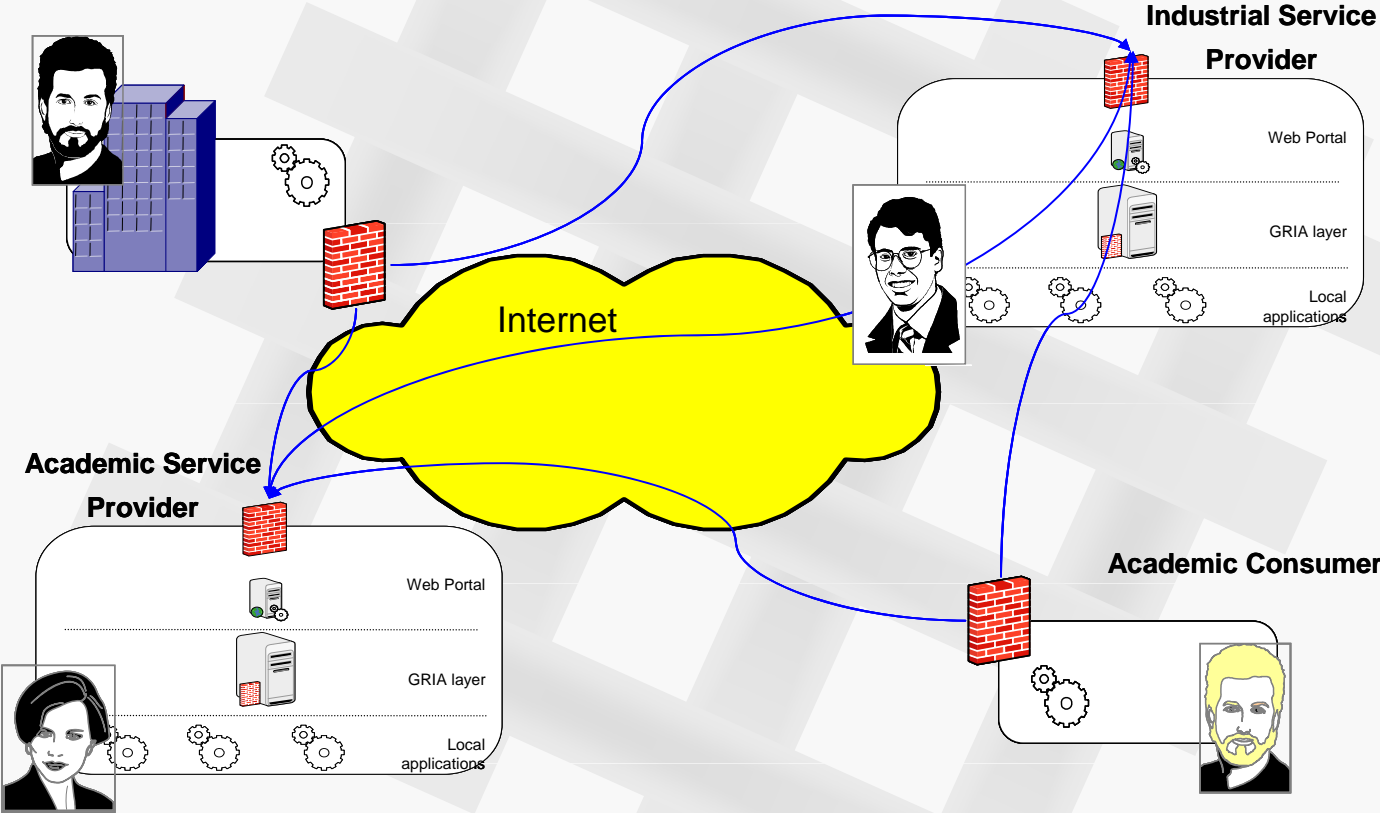
- Develop generic Grid technology for the solution of complex application problems
- Use this new technology in the targeted industrial application sectors
- **Enhance Data Grid technology** for product development and production process design
- **Raise awareness** of the advantages of Data Grids in important industrial sectors

Project Organization



GSK, Inpharmatica, ULB, FhG-SCAI-Bio,UKA, EMBL, InforSense, and NEC

Business to Academia Prototype (B2A)



Targeted pharmaceutical Workflow(s)

- Master Sequence Analysis Pipeline
 - Definition and Design by GSK
 - MSAP will initially focus on **decorating sequence data** with annotation from sources designed to better validate sequence structure and function as well as annotation species availability
 - Get high quality state of the art analysis for GSK genes of interest
 - These analyses are currently confined to systems within GSK
 - **Broaden the scope** of their analysis and be able to import the best of breed analysis from both Academia and Vendors at costs appropriate to the type of analysis
- (IXodus Workflow)
 - Characterisations of genes expressed in the salivary gland of the tick ***Ixodes ricinus***

MSAP - Client

The screenshot displays the InforSense KDE interface. On the left, a 'Resources' pane shows a tree view of 'Userspace' containing various tasks and components. The main workspace shows a 'Data-flow' diagram with nodes such as 'Input Accession', 'Get Sequence', 'Run Antigenic (ULB)', 'Data Resource', 'Get Expression', 'Annotate Expression', 'Get SNPs', 'Annotate SNPs', 'Get Transcripts', 'Annotate Transcripts', 'Get Annotation', and 'Annotate Other'. A 'Deployable FeatureVista' node is also present. Below the diagram is the 'Node Editor [Add Annotation or Feature] Annotate Expression' window, which includes tabs for Parameters, Output, Cache, History, Notes, Interaction, and Input. A table in the editor shows the following configuration:

Name	Value
Type	feature
Sequence ID	ACCESSION
Feature Type	FEATURE TYPE
Feature Source	FEATURE SOURCE
Feature Start	sts start int

At the bottom of the node editor, a warning icon is accompanied by the text 'Build cache or execute before proceeding'. The status bar at the bottom right shows 'Tasks pending: 0 | Tasks running: 0'.

SIM



Remote Academic Services - ULB

The screenshot displays the InforSense KDE interface, a workflow management tool. The main window shows a data-flow diagram with several interconnected tasks:

- test.seq** (red cube icon) connects to **Group[...Read File]** (blue cube icon).
- test.seq (2)** (red cube icon) connects to **Run Primer Design** (blue cube icon).
- Run Primer Design** connects to **Primer Design Results** (red cube icon).
- seq** (red cube icon) connects to **Upload File** (purple cube icon).
- Upload File** connects to **Extract Data** (purple cube icon).
- Extract Data** branches into two paths:
 - One path goes to **Run Antigenic Site Finding** (blue cube icon), which then connects to **Download Result** (purple cube icon), and finally to **Antigenic Results (Orch)** (red cube icon).
 - The other path goes to another **Download Result** (purple cube icon), which connects to **Protein Results** (red cube icon).

The interface includes a menu bar (File, Edit, Resources, Workflow, Notebook, View, Tools, Help), a toolbar, and a left-hand sidebar with a tree view of resources. The bottom status bar shows "3 nested warnings" and "Tasks pending: 0 | Tasks running: 0".

SIM



Technology Uptake

- ***Integrated Grid Infrastructure***
 - GRIA (IT Innovation)
- ***Workflow Design and Enactment***
 - KDE (InforSense)
- ***Data Integration Tool***
 - MRS (CMBI University of Nijmegen)
- ***Semantic Service Registry***
 - Semantic Broker (NEC)
- ***Service Annotation***
 - TUAM (FhG-SCAI-BIO)
- ***P2P File System***
 - IGOR (University of Karlsruhe)

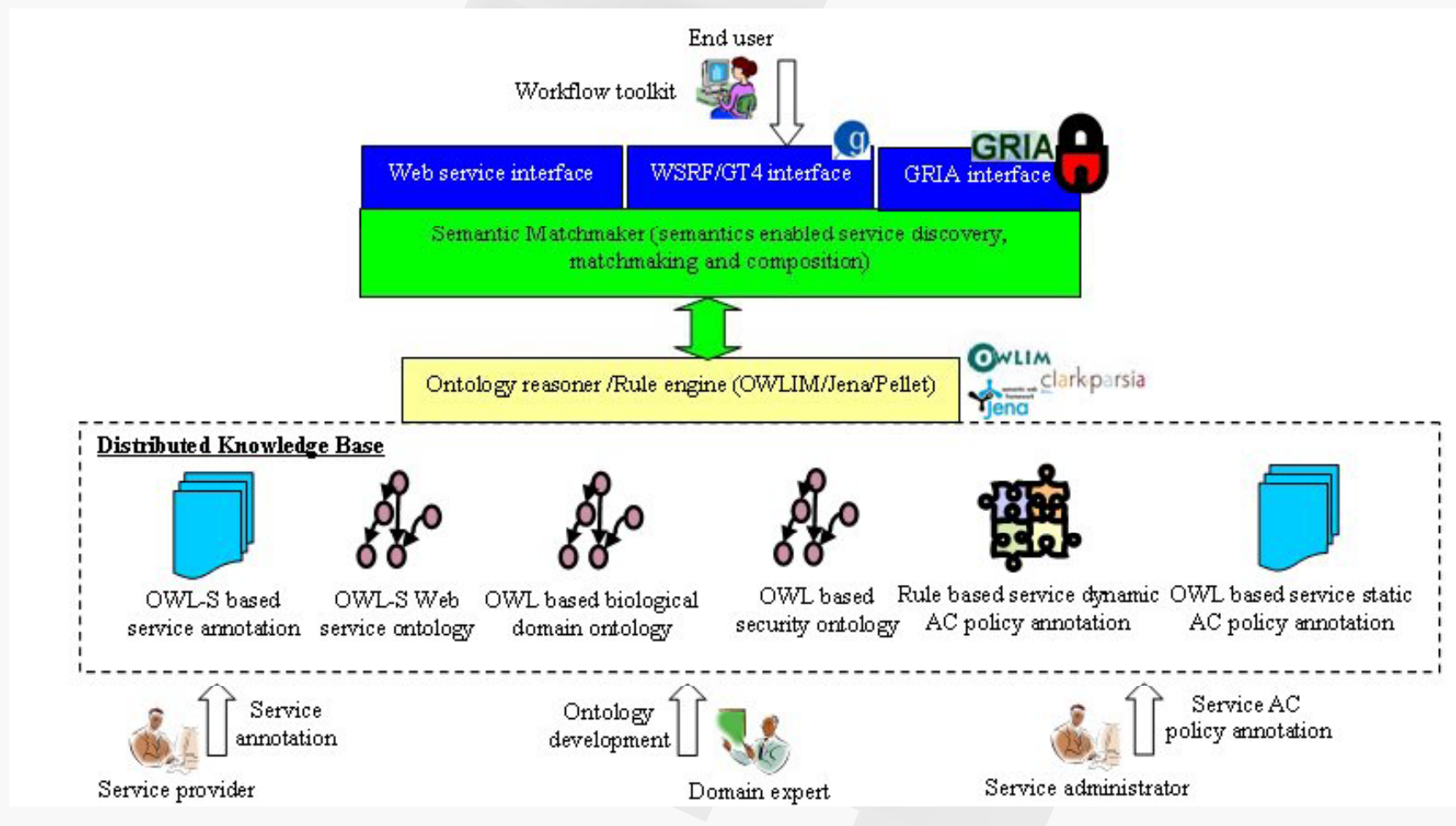
GRIA

- Developed in IST project (2001-04)
- Make the Grid usable by industry
- Bases on Web Service technology
- Access only on pre-defined services
- Authorisation via process-based access control
- Integration of an flexible SLA system
- Current version: GRIAv5.1

KDE

- Workflow tools are mainstream in the Life Science sector
 - KDE licenses available at industrial partners before the project
 - Academic user experienced with Taverna
- Development of Grid nodes
 - GRIA
- Creation of Semantic Broker node

Semantic Broker Architecture



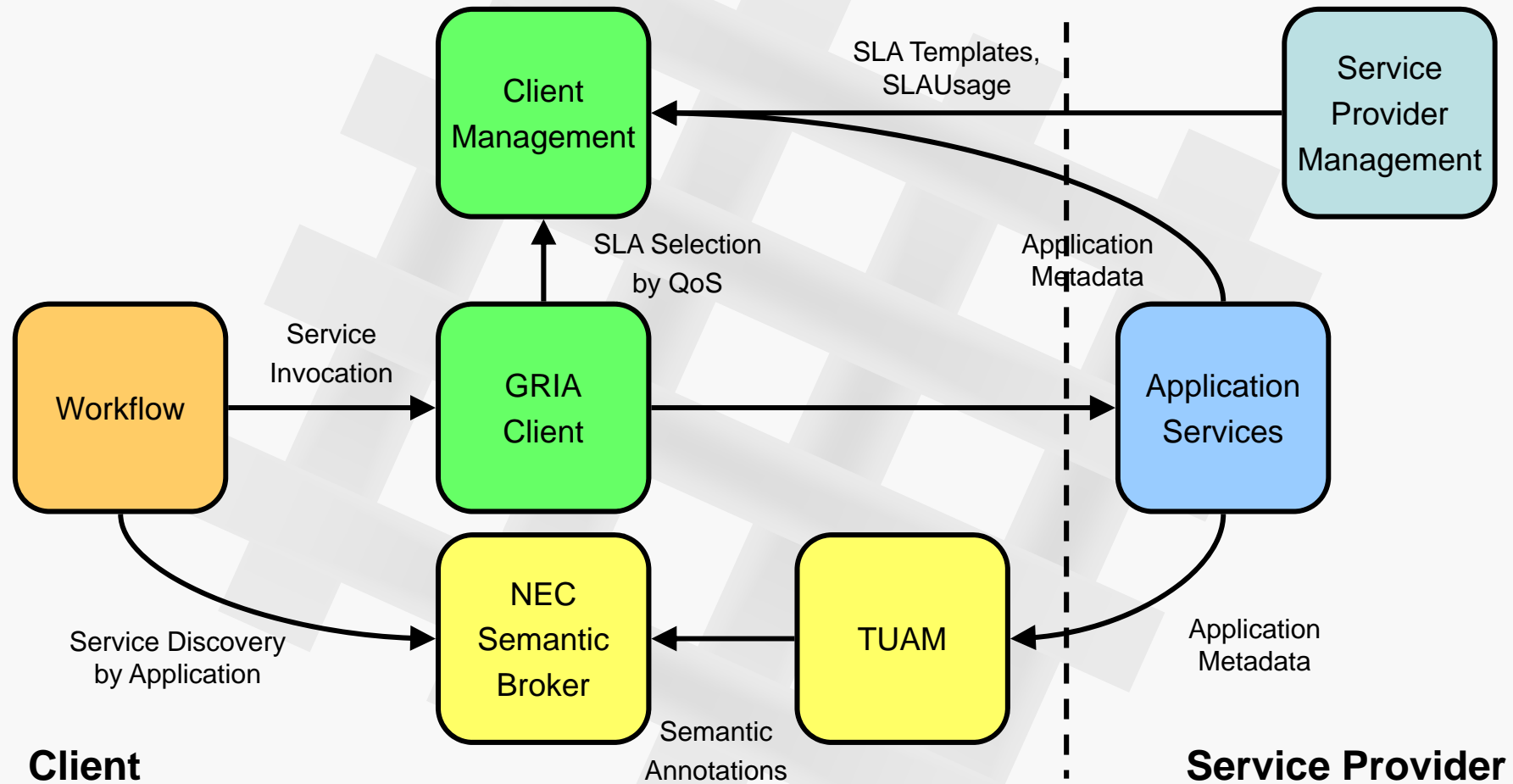
Semantic Broker Key Technologies

- OWL-DLP/OWL-DL based biological domain ontology and security ontology
- OWL-S based service annotation
- Semantic matchmaking based on ontology reasoning
- Semantic annotation of service AC policies using ontologies and semantic rules
- Controlled discovery of services based on the user credential and service AC policy annotation.

Semantic Broker Key Features

- Standard oriented design and implementation (OWL, OWL-S, SWRL)
- Fully based on open source products (OWLIM/Sesame, Jena, Pellet, OWL-S editor/SWRL editor/Protégé)
- Interoperable with other OWL based bioinformatics ontologies (e.g., GO, SO)
- Deployable within mainstream Grid middleware (GT4, GRIA, Web service), Secure access supported through GRIA
- Supported by Bioinformatics specific service annotation/publication tools (TUAM, Dynamo)

Discovery and Selection Infrastructure



Courtesy of IT Innovation

Success Criteria

- **Usable** by scientists within industrial organisation with minimal IT support
- **Flexible** that they may be able to change and scale to the changing environment
- **Capable** to lower the barriers for working with external partners in a virtual organisation
- **Compliant** with industry's stringent security policies
- **Responsive** enough to be usable in a timely fashion

Impact on industrial Partners (1)

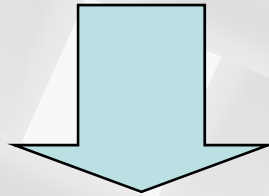
- Lowering the barrier of building a business relationship
- Money (e.g. lawyer fees)
- Knowledge
- ...

Scale the relationship!



Impact on industrial partners (2)

- Access to new markets
- Finer granularity of the product portfolio
- Increase the commercial offer
- Faster deployment



New Business Model

Role of Semantic Technologies

- Semantic service discovery
 - Annotation of services (TUAM)
 - Semantic Registry (Semantic Broker)
- Industrial partners recognise them still as research topic
 - Uptake in very near future unlikely
- Academic users / domain experts are already using it now
- Potential to ease the exploitation of the growing number of resources

Conclusions

- Grid technology seems to play an important role in life science
- Common standards and agreements are missing
- Traceability of activities from the industrial have to sufficiently obfuscated
- Experiences have to be made with low risk data

Thank you

Questions ?