An Introduction to SIMDAT
a Proposal for an Integrated Project on EU FP6 Topic

“Grids for Integrated Problem Solving Environments”

Martin Hofmann

Department of Bioinformatics

Fraunhofer Institute for Algorithms
and Scientific Computing (SCAI)

“Grids for Complex Problem Solving“
a Strategic Objective of the Framework 6 IST Programme

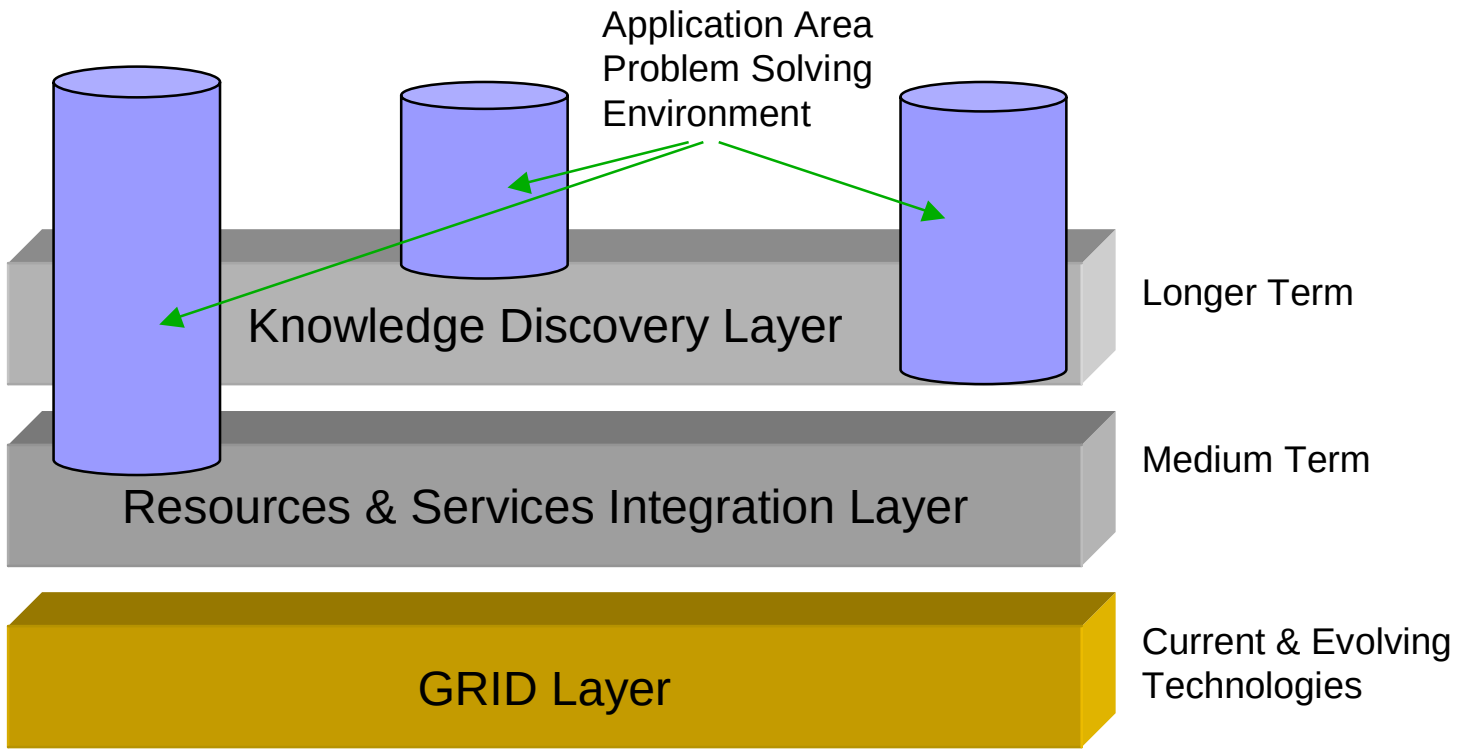
What is different to Grid in FP 5 ?

- focus on „next generation GRID“
- focus on enabling application technologies
- focus on industrial applications rather than academic projects
- focus on generic approaches that can be used in different application areas

Fraunhofer Institut Algorithmen und Wissenschaftliches Rechnen

EC Grid Workshop on GRID-based Complex Problem Solving Environments

held at Schloss Birlinghoven, St. Augustin, Germany, on April 29th to 30th, 2003



Fraunhofer Institut Algorithmen und Wissenschaftliches Rechnen

Lessons from the Birlinghoven Workshop

A) data integration in the GRID context is still a largely unsolved problem

B) data integration in the GRID context faces the same problems as data integration in non-GRID context, namely the need for “semantic glue“

C) limitations to generic GRID approaches in different industrial application areas result from different semantics in those industrial sectors

D) representations of objects and semantics by ontologies are key to interoperability and knowledge discovery

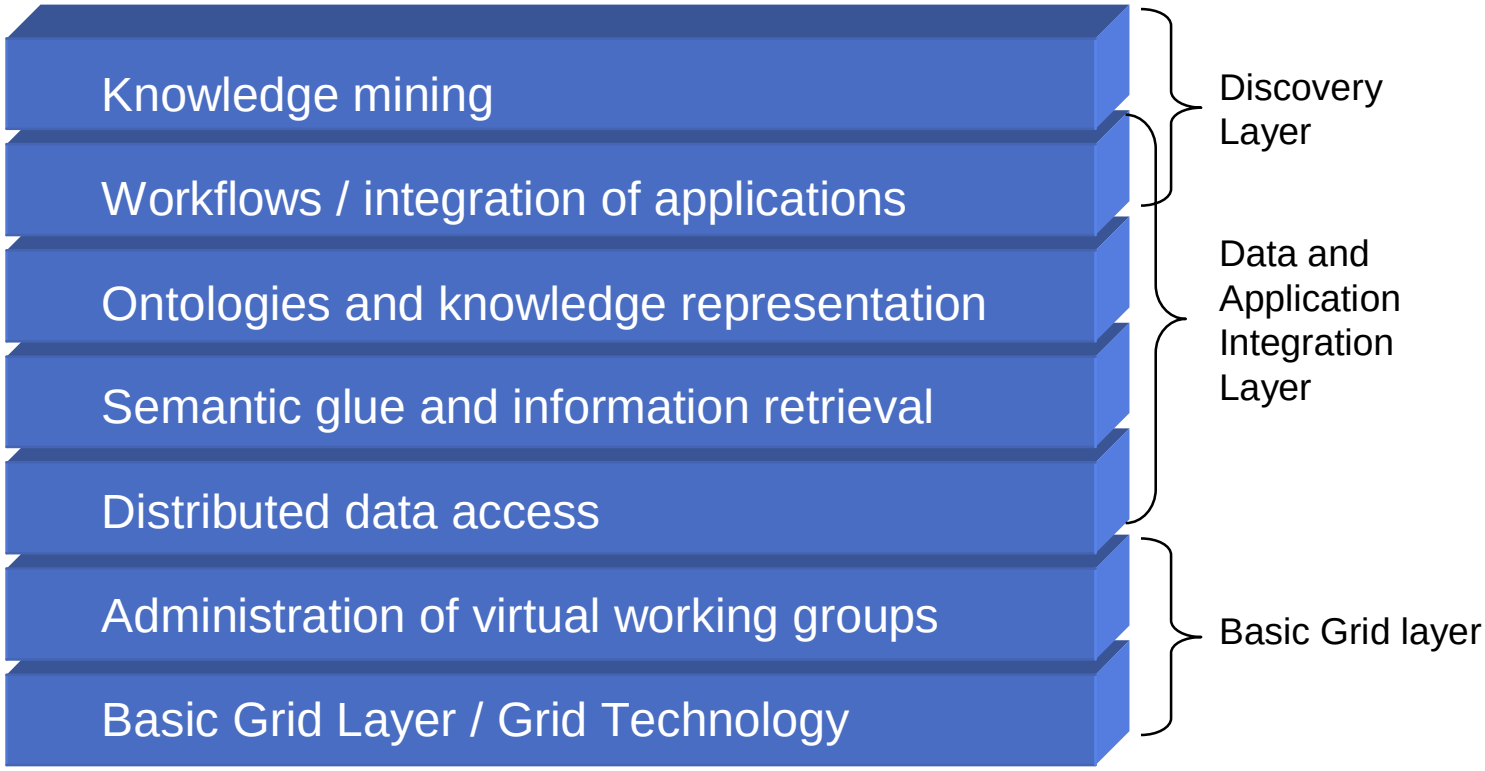
Fraunhofer Institut Algorithmen und Wissenschaftliches Rechnen

…. more Lessons from the Birlinghoven Workshop

Most important: the notion is that issues like semantic integration ("semantic glue") and the representation of objects and their relationships by means of ontologies is most advanced in the field of bioinformatics.

Therefore, bioinformatics will play a key role in demonstration activities that aim at implementing a first GRID-based complex problem solving environment.

Fraunhofer SCAI Institut Algorithmen und Wissenschaftliches Rechnen



Fraunhofer Institut Algorithmen und Wissenschaftliches Rechnen

SIMDAT: Proposal for an Integrated Project in FP6

Objectives:

To expand the potential of the GRID approach to solving complex problems which can not be solved with current technologies in application fields such as, but not limited to, industrial design, engineering and manufacturing, **health, genomics and drug design**, environment, critical infrastructures, energy, business and finance, and new media.

To overcome present architectural and design limitations hampering the use and wider deployment of computing and knowledge GRIDs and to enrich its capabilities by including new functionalities required for complex problem solving. This should help the larger uptake of GRID type architectures and extend the concept **from computation Grids to knowledge GRIDs, eventually leading to a "semantic GRID".**

Fraunhofer Institut Algorithmen und Wissenschaftliches Rechnen

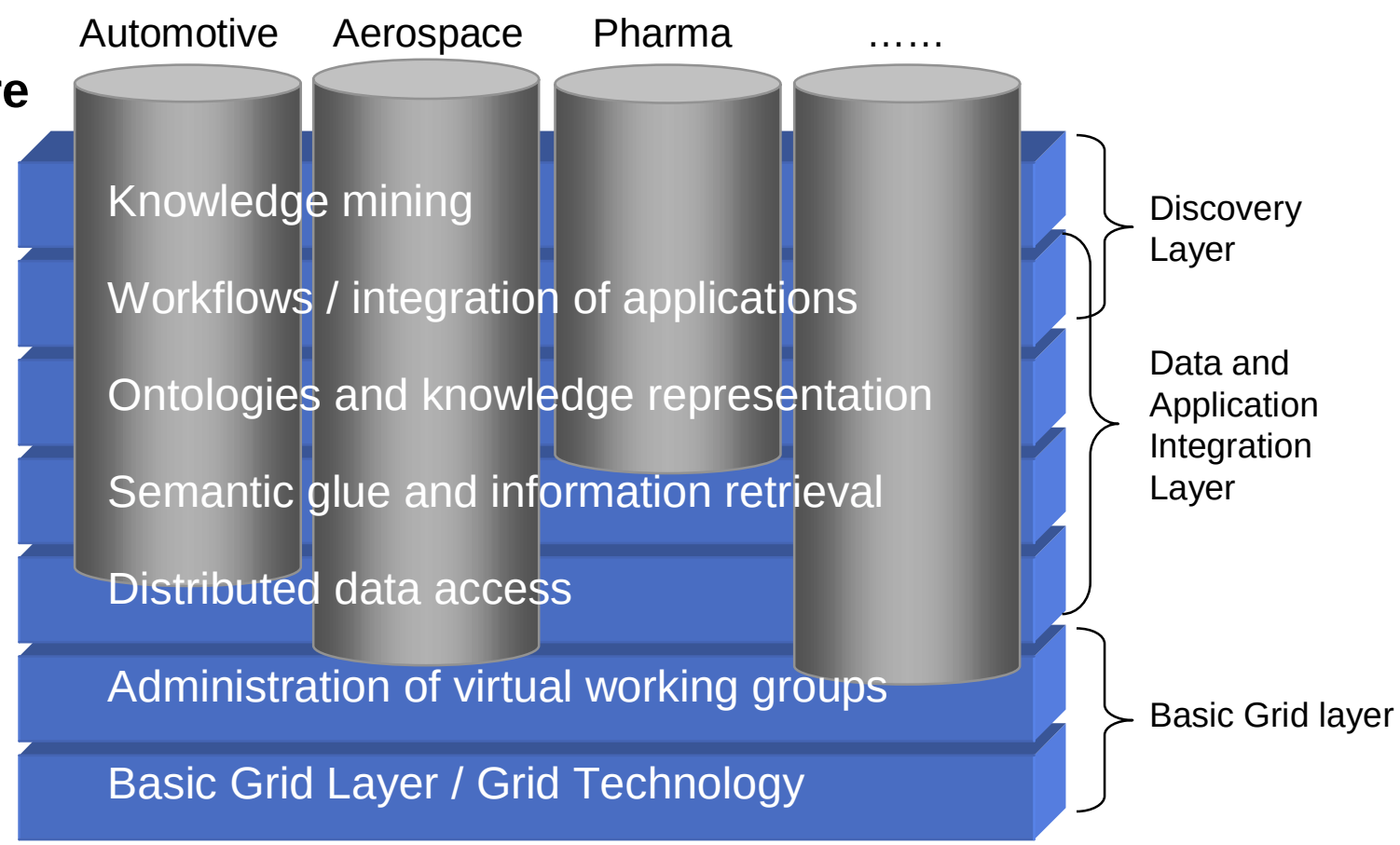
Application Areas

Industrial Sectors

- Automotive
- Pharma
- Chemical Engineering
- Oil Industry
- Aircraft Design
- Satellites
- Climate

Fraunhofer Institut Algorithmen und Wissenschaftliches Rechnen

SIMDAT Structure



Fraunhofer Institut Algorithmen und Wissenschaftliches Rechnen

Bioinformatics and the GRID

Fraunhofer Institut
Algorithmen und Wissen-
schaftliches Rechnen

A politically almost incorrect statement ….

It seems that with GRID technology you have all problems that you already had with “traditional” Server-Client or Web-based approaches *PLUS* significant additional effort on defining standards, generating trust and monitoring / documenting access to resources and tools.

Let alone the implications on licensing ….

Fraunhofer Institut Algorithmen und Wissenschaftliches Rechnen

GRID-based Complex Problem Solving and Bioinformatics (I)

A GRID – based complex problem solving environment in bioinformatics should:

- seamless integrate existing data and databases
- re-use as much as possible from existing infrastructures
- make the advantage of using GRID technology *immediately apparent* to the user community
- align with existing and emerging initiatives in the field of knowledge representation / ontologies

Fraunhofer Institut Algorithmen und Wissenschaftliches Rechnen

GRID-based Complex Problem Solving and Bioinformatics (II)

A GRID – based complex problem solving environments in bioinformatics should therefore:

- be limited to an existing academic or industrial organisation
- this organisation should be a global organisation
- there has to be trust amongst the participating units within the organisation; IP issues should be clear

Fraunhofer Institut Algorithmen und Wissenschaftliches Rechnen

The SIMDAT Bioinformatics Cluster

Currently, a consortium of public research institutions, large pharma companies and SMEs is assembled to form the bioinformatics cluster within SIMDAT. As outlined above, the goal is to expand the potential of the GRID approach to solving complex problems. We intend to do this by implementing working solutions for information retrieval, domain-specific querying and knowledge mining based on the existing, basic GRID technology and data access protocols.

The SIMDAT Bioinformatics Cluster: Tasks

The SIMDAT bioinformatics cluster will focus on challenges associated with different levels of the SIMDAT structure. In particular, we will address the following issues:

- Evaluation of existing (GRID) information retrieval systems and adaptation to the needs of life science informatics
- “warehousing” of controlled vocabularies and ontologies and development of a biomedical GRID annotation service to facilitate annotation of phenotypes, experimental conditions, and images
- New ways to use these annotations to establish semantic links between very heterogeneous data types (e.g. histological images and relevant disease genes that are expressed in the same tissue; knowledge mining based on annotation paths)

Fraunhofer Institut Algorithmen und Wissenschaftliches Rechnen

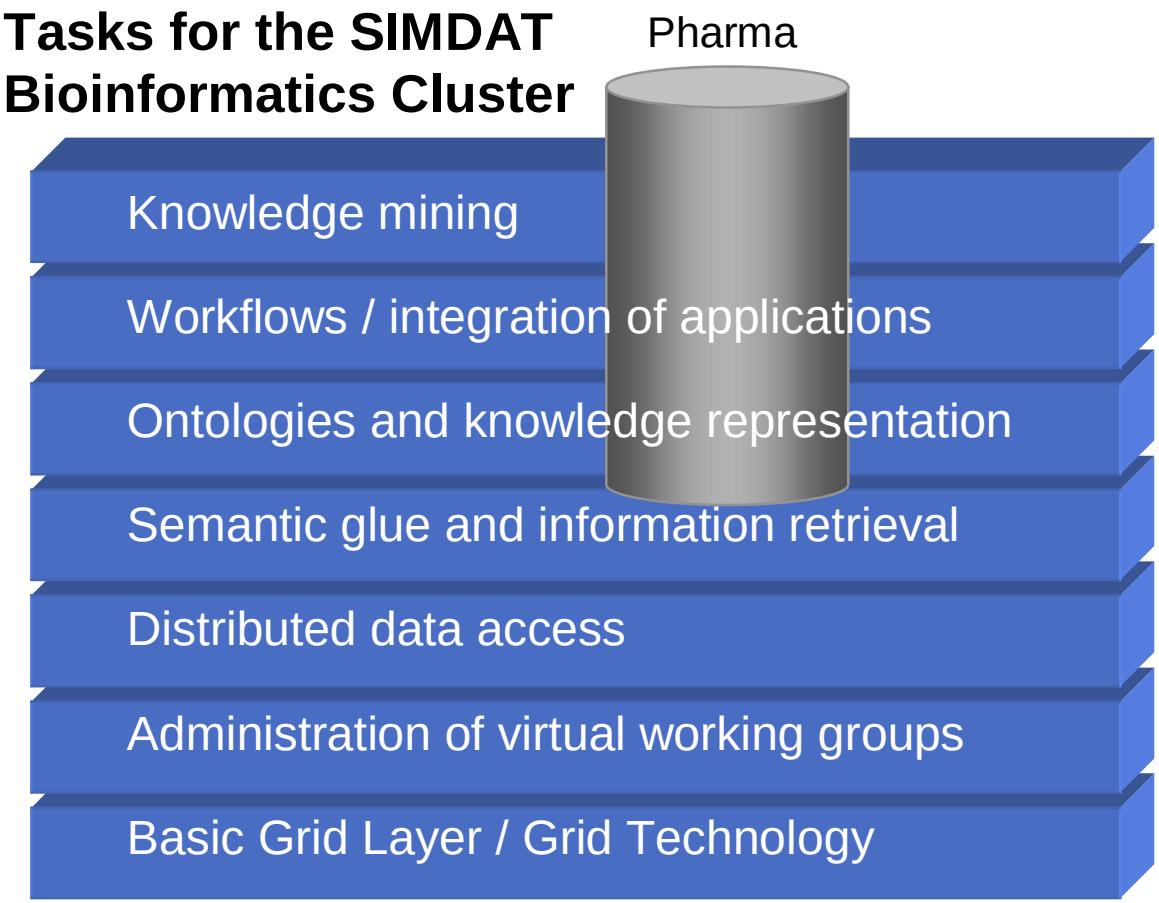
The SIMDAT Bioinformatics Cluster: Application Scenarios

Currently, we have three different application scenarios in our mind:

- A GRID-based complex problem solving environment for an existing, successful, global academic consortium → EMBnet
- A GRID-based complex problem solving environment supporting the partnering of biotechnology industry with big pharma (a B-2-B scenario) → ACE bioscience
- A GRID-based complex problem solving environment for a large, global pharmaceutical research organisation → GSK

Fraunhofer Institut Algorithmen und Wissenschaftliches Rechnen

Tasks for the SIMDAT Bioinformatics Cluster

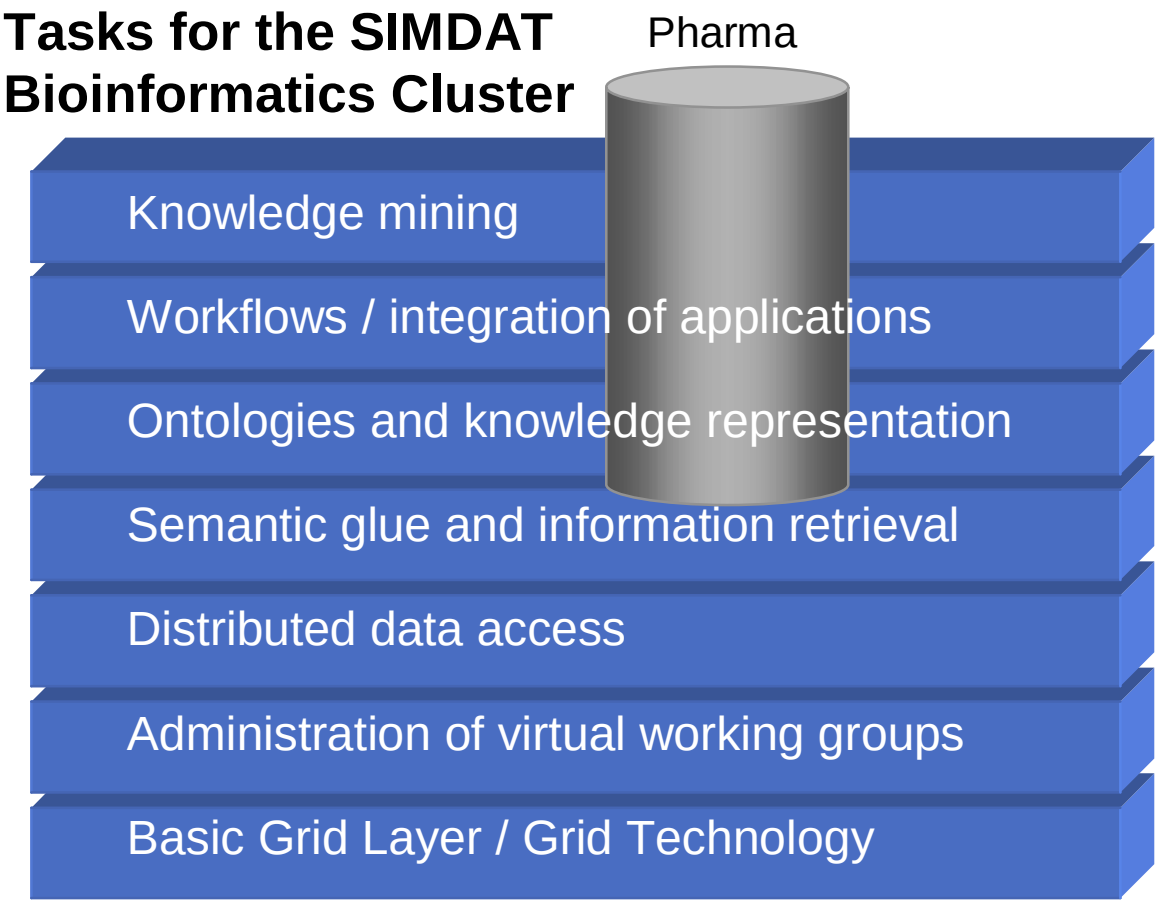


Evaluation of existing (GRID) information retrieval systems and adaptation to the needs of life science informatics

→DiscoveryLink ?
→federated SRS ?
→XML-based mediators ?
→ …..

Fraunhofer Institut Algorithmen und Wissenschaftliches Rechnen

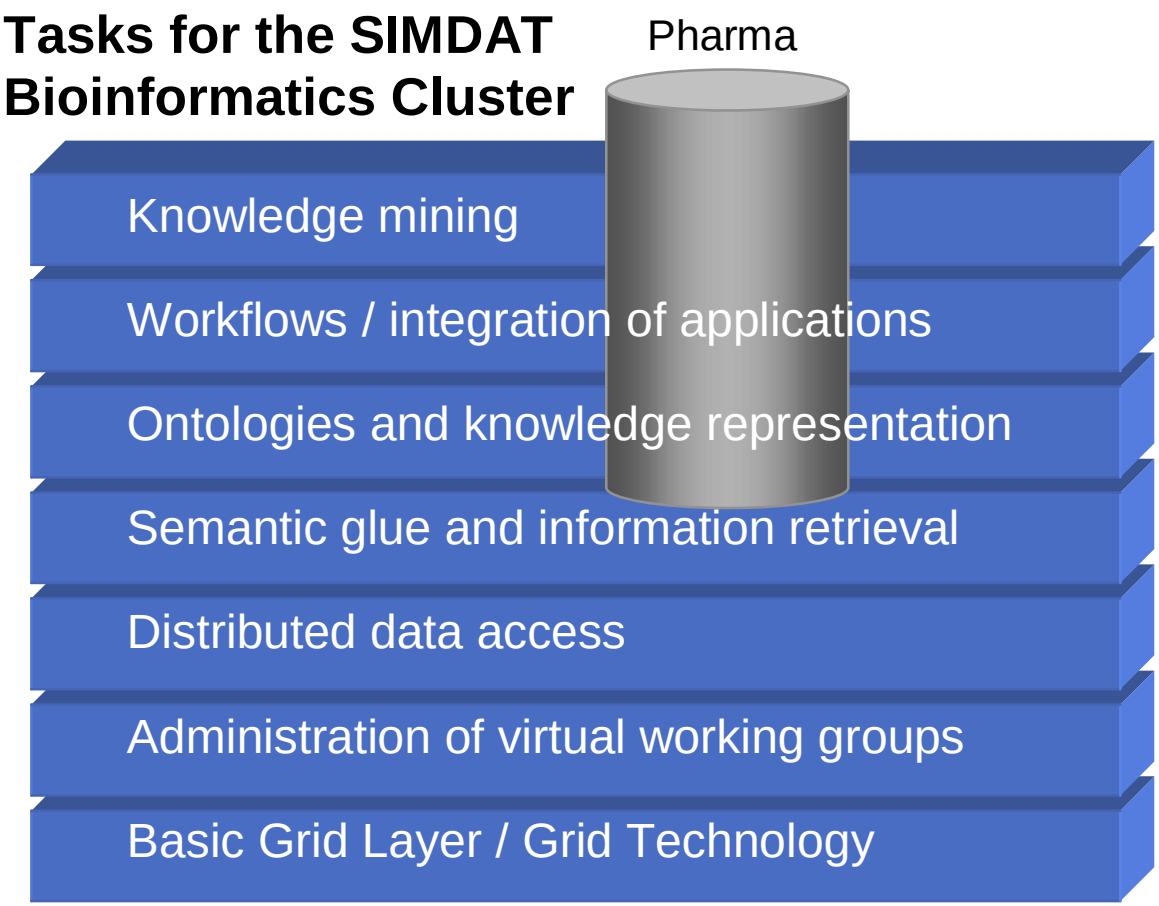
Tasks for the SIMDAT Bioinformatics Cluster



"warehousing" of controlled vocabularies and ontologies and development of a biomedical GRID annotation service to facilitate annotation of phenotypes, experimental conditions, and images

Fraunhofer Institut Algorithmen und Wissenschaftliches Rechnen

Tasks for the SIMDAT Bioinformatics Cluster



"New ways to use these annotations to establish semantic links between very heterogeneous data types (e.g. histological images and relevant disease genes that are expressed in the same tissue; knowledge mining based on annotation paths)

Fraunhofer SCAI Institut Algorithmen und Wissenschaftliches Rechnen

The SIMDAT Bioinformatics Cluster: Partners

The following partners are in discussions on participation in SIMDAT/bio:

Thure Etzold	LION bioscience Ltd. (UK)	SME
Robert Herzog	EMBnet.org (B)	academic consortium
Rob Gill	Glaxo Smith Kline (UK)	Pharma
Martin Hofmann	Fraunhofer SCAI (GER)	Academic institution
Angus King	ACE bioscience (DK)	SME
Guy Lonsdale	C&C Research Laboratories (UK)	IT technology
Thomas Fuhrmann	University of Karlsruhe (GER)	Academic institution
Jarek Nabrzyski	Poznan Supercomp. Centre (Po)	Academic institution
Piotr Zielenkiwicz	Poznan EMBnet node	
……	……	……

Fraunhofer Institut Algorithmen und Wissenschaftliches Rechnen

The SIMDAT Bioinformatics Cluster: Task I

Evaluation of existing (GRID) information retrieval systems and adaptation to the needs of life science informatics → one road to go: federation of SRS

Talk continued by Thure Etzold, LION bioscience Ltd., Cambridge (UK)

Fraunhofer Institut Algorithmen und Wissenschaftliches Rechnen