Overview of the e-LICO Project

e-LICO Kick-Off Meeting

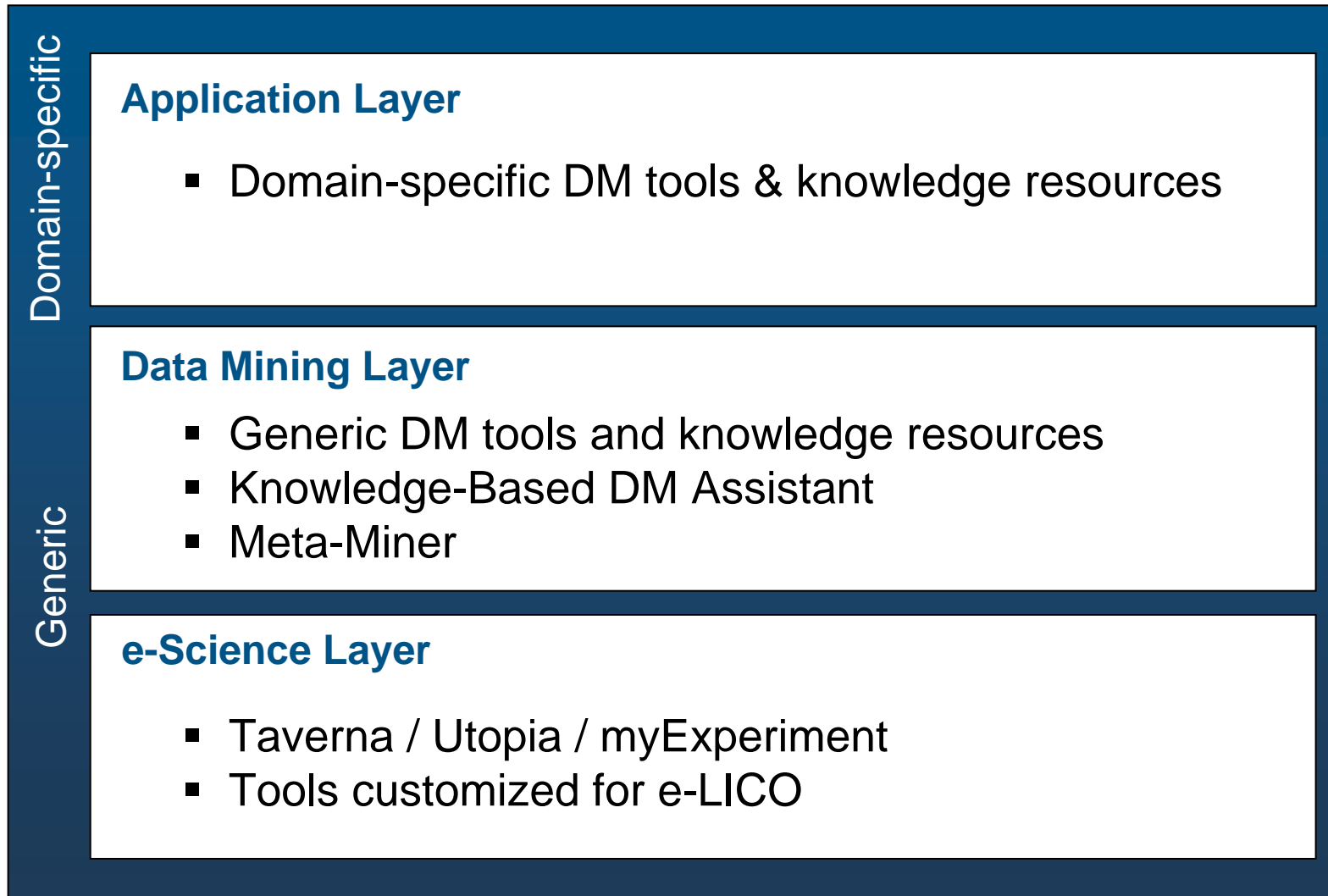Geneva, February 12-13, 2009

Main project deliverable

A virtual data mining laboratory that features

- a collaborative e-science infrastructure

- a self-improving DM assistant
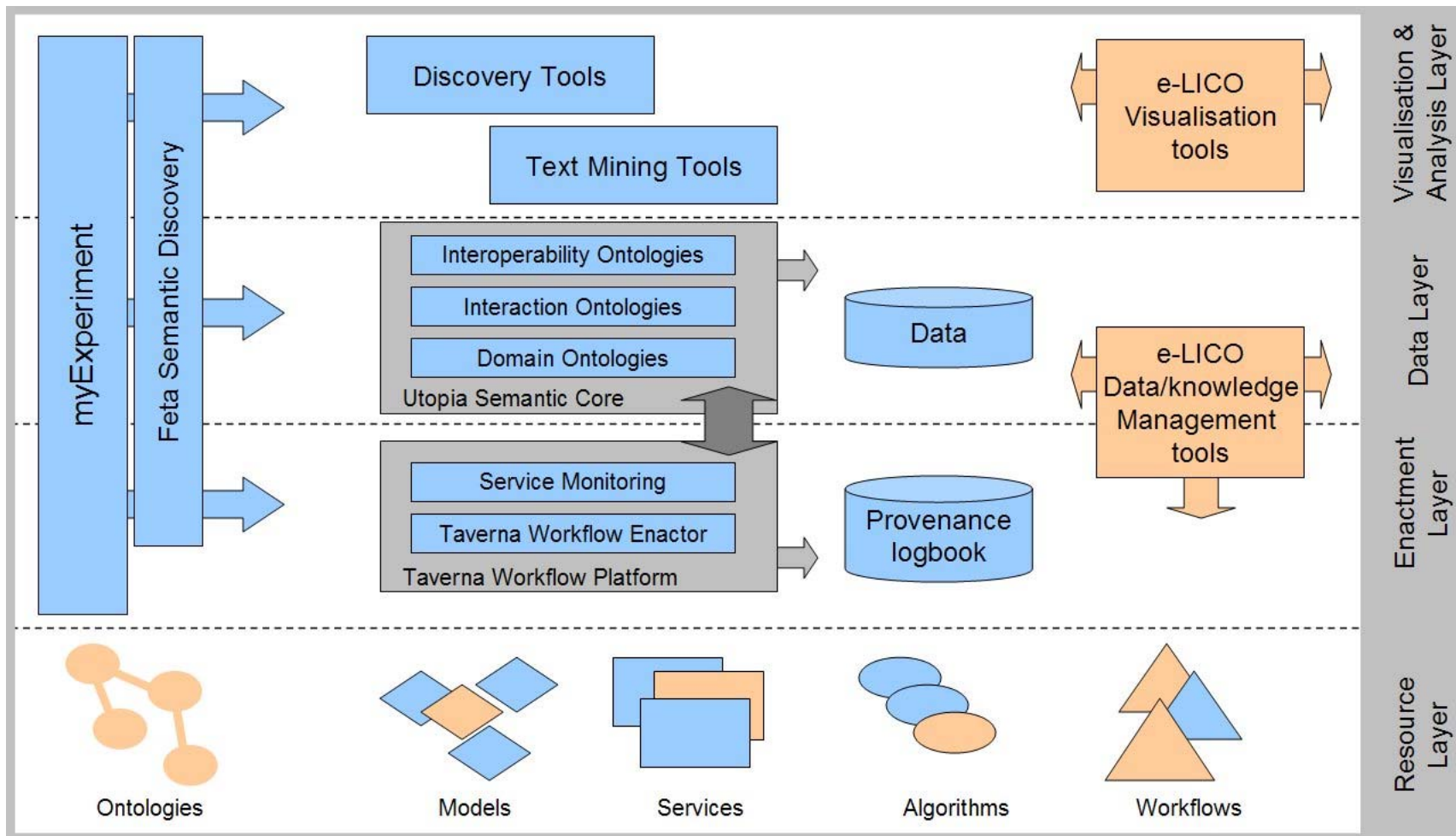
- adaptability to different sciences

Key ideas

- ontology-based planning to build DM workflows

- incremental self-improvement through meta-mining

- a novel meta-learning technique that blends

    - probabilistic reasoning

    - kernel-based learning from complex structures

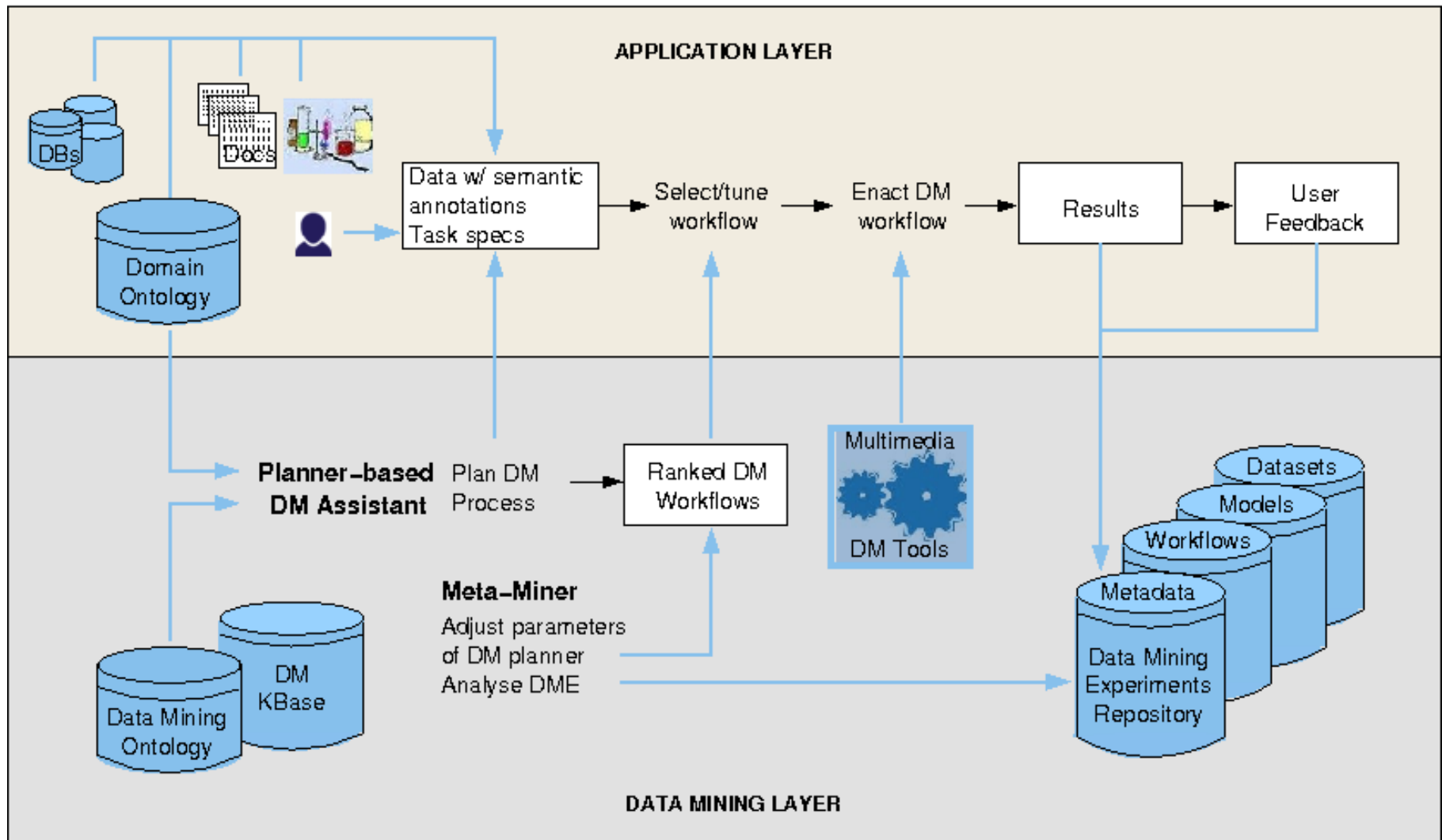- reliance on social networks of committed scientists

System Architecture

Domain-specific

Generic

Application Layer

- Domain-specific DM tools & knowledge resources

Data Mining Layer

- Generic DM tools and knowledge resources
- Knowledge-Based DM Assistant
- Meta-Miner

e-Science Layer

- Taverna / Utopia / myExperiment
- Tools customized for e-LICO

The DM and application layers

The pilot application

- The application layer is an empty shell

- Each instance of the DM lab comes from filling this shell

- Pilot domain: systems biology approach to the kidney and urinary pathways (KUP)

- Close links with COST Action EuroKUP (2008-2011)

Concrete project outputs

- **Major output: generic e-LICO**

  - Generic software

  - Generic DM knowledge sources

- **By-product: KUP application layer**

  - Biological software

  - Biological knowledge resources

Generic Products

Software

- e-LICO infrastructure and tools for collaborative authoring and annotation

- DM, TM and IM tools developed in the project

- DP, DM, TM and IM workflows

- Intelligent discovery assistant

- Probabilistic, kernel-based meta-miner

Content

- DM ontology

- DM KDB

- DM Experiment Repository (potential counterpart of UCI ML Repository for meta-data)

Domain-specific by-products

Software

- Biological data, text and image mining tools

Content

- KUP ontology
- KUP knowledge/data base
- Prédictive models for KUP diseases
- Repository of KUP data mining experiments