



**e-LICO is a virtual laboratory** for interdisciplinary collaborative research in data mining and data-intensive sciences. e-LICO consists of three layers: the e-science and data mining layers form a generic research environment that can be easily adapted to different scientific domains by customizing the application layer.

**Partners:**

- University of Geneva (Switzerland)
- Institut national de la santé et de la recherche médicale (France)
- Medice Oy (Finland)
- National Hellenic Research Foundation (Greece)
- Rapid-I GmbH (Germany)
- University of Manchester (UK)
- University of Zurich (Switzerland)

**Application layer (domain specific)**

- Customized tools/services/workflows for mining domain-specific data
- Domain-specific knowledge resources (specialized ontologies, knowledge and data bases)

**Data mining layer (generic)**

- Tools/services for multimedia data mining
- Data mining knowledge resources (ontology, knowledge/data bases)
- Knowledge-driven data mining assistant
- Meta-miner for improvement of data mining workflows

**E-science layer (generic)**

- E-science infrastructure extended with mechanisms for multi-level, multi-modal collaboration
- E-science infrastructure extended with tools and workflows for knowledge acquisition/management, e.g. ontology engineering, semantic annotation

*Collective Intelligence*

# Data Mining

## e-Science Laboratory



**The e-science layer**, built on an open-source e-science platform, supports content creation through collaboration at multiple scales and degrees of commitment – ranging from small, contract-bound teams to voluntary, constraint-free participation in dynamic virtual communities.

**The data mining layer** is the distinctive core of e-LICO; it provides comprehensive access to multimedia data mining tools. Standard tools are augmented with preprocessing or learning algorithms developed specifically to meet challenges of data-intensive, knowledge rich sciences.

**The virtual lab** enables scientists out of different areas to collaborate. For example, data mining practitioners help biologists to scan their data for unknown patterns. And biologists pose new challenges for the development of highly scalable algorithms. The support of this collaboration is one of the crucial aspects of the e-LICO project.

**Methodologically sound use of these tools** is ensured by a knowledge-driven data mining assistant, which relies on a data mining ontology and knowledge base to plan the mining process and propose ranked workflows for a given application problem. The collaborative work in the e-lab is continuously monitored and hence experimental meta-data is collected in order to support replication and comparison of data mining experiments. This meta-data is used by a meta-miner, which will combine probabilistic reasoning with kernel-based learning from complex structures to incrementally improve the assistant's workflow recommendations.

**The ultimate result** of this combination of collaborative work and our meta mining approach is a recommendation system for non-experts in data analysis: “scientists with a similar problem like yours tried the following analysis processes...”.

**e-LICO is showcased** in a systems biology task: biomarker discovery and molecular pathway modelling for diseases affecting the kidney and urinary pathways. The final deliverable of the project will be a free, experimental prototype open to continuous collaborative expansion and refinement by the research community.