

## AutoMed: Automatic Generation of Mediator Tools for Heterogeneous Data Integration



### Project Leader

Alex Poulouvassilis

### Project staff

Lucas Zamboulis

Sandeep Mittal

Jianing Wang

Dean Williams

### Other Project Partners

Imperial College

### Project Details

2001 to date

Funded by EPSRC, MoD, BBSRC

### Project Web Site

Please see here for further details of the project, downloadable software and articles:

[www.doc.ic.ac.uk/automed](http://www.doc.ic.ac.uk/automed)

### Keywords

Data Transformation

Data Integration

Metadata Management

### Project Aims

The AutoMed project is developing tools to assist users and developers in the transformation and integration of data from different data sources. The AutoMed toolkit can access structured (e.g. relational), semi-structured (e.g. XML, RDF) and text data sources. Data transformation/integration is achieved by defining **transformation pathways** between schemas, for example between a set of data source schemas and a 'virtual' Integrated Schema, as illustrated on the right. The AutoMed toolkit comprises a number of components, and at the LKL we are currently working on the following tools:

### Query Processing

The AutoMed Query Processor answers queries expressed over an Integrated Schema. It uses the transformation pathways between the Integrated Schema and the data source schemas to reformulate the query, optimizes the reformulated query, submits sub-queries to the data source wrappers for evaluation, and finally merges the sub-query results into an overall result for the original query.

### XML Transformation & Integration

This tool provides facilities for transforming/integrating XML data. It can extract a schema from an XML file, use schema-matching techniques to reconcile different XML schemas, restructure a set of reconciled schemas into a single integrated schema, and then use this integrated schema for querying the original XML files or creating one integrated file.

### Schema analysis and auto-reconstruction

This tool assists in the semi-automatic creation of a global schema and mappings for a given set of data sources. After the integrator creates a construct in the global schema, the tool determines the correctness of this construct and of the

associated mappings using formalisms generated from user-defined requirements and suggests possible changes.

### Text Information Extraction Tool (ESTEST)

This tool generates annotations on text sources using grammar rules. New data and metadata are then derived from these annotations which can be integrated with other data sources via AutoMed transformation pathways.

AutoMed is currently being used to support heterogeneous biological data integration in the ISPIDER and BioMap projects and medical data integration in the ASSIST project.

