Research Projects under FP6 and FP7: INFOBIOMED, ACGT & ACTION-Grid

Strategies for knowledge discovery, management and integration in multi scale and heterogeneous biomedical data sources

Diana de la Iglesia
Biomedical Informatics Group
School of Computer Science
Universidad Politécnica de Madrid
diglesia@infomed.dia.fi.upm.es

The Action-Grid project is funded by the European Commission under the 7th framework program.
INDEX

• Introduction
• Knowledge Discovery in Databases & Data Integration
• INFOBIOMED: Network of Excellence
• ACGT: Integrated Project
• ACTION-Grid: Support Action
The Biomedical Informatics Group (GIB) is involved in several research projects related to:

- Knowledge discovery
- Data mining and text mining
- Management and integration in multiscale and heterogeneous biomedical data sources
- Services for heterogeneous biomedical data integration
- Nanomedicine
- Grid computing

Experience in EC projects:

- Projects under FP5: INFOGENMED
- Projects under FP6: INFOBIOMED and ACGT
- Projects under FP7: ACTION-Grid
Knowledge Discovery in Databases (KDD)

The Action-Grid project is funded by the European Commission under the FP7
Integration of schemas

Database → Physical Schema → “Mapping” relations → Conceptual Schema → Domain Ontology

Table_1
Attribute_1
...  
Attribute_n

Table_2
Attribute_1
Attribute_2
...  
Attribute_n

Table_n
Attribute_1
Attribute_2
...  
Attribute_n

Concept_1
Rel_1
Concept_2
Rel_2
Concept_3
Rel_3

...  

Concept_n
Rel_n

The Action-Grid project is funded by the European Commission under the FP7
Data inconsistencies

Pre-processing
- Data source
  - Data source
  - URL
- Pre-processing
- URL
- URL

Data transformation
- Order
  - Order
  - Target fields
- Condition
  - Condition
  - Representative Values
  - Values Ranges
- Conditions
- Duplicates
  - Duplicates
  - Mathematical Expression
  - Mathematical Expression
- Missing Values
  - Missing Values
  - Representative Values
  - Representative Values
- Transformations
  - Transformations
  - Fixed Value Substitution
  - Most Frequent Value Substitution
  - Average Substitution
  - Row deletion
- Detectors
  - Detectors
  - Values Ranges
  - Values Ranges
- Patterns
  - Patterns
  - String
  - Regular expression
  - Adjustment rule
  - Substitution
- Synonyms
  - Synonyms
  - Favourite term
  - String
- Target fields
  - Target fields
Method to detect instance inconsistencies

1. Tables to analyse
2. Pre-processing ontologies
3. Analysing data sources
4. Identification of inconsistencies
5. Confidence level
6. Generation of suggestions
7. Supervision by an expert

The Action-Grid project is funded by the European Commission under the FP7
INFOBIOMED

Structuring Biomedical Informatics to support individualised healthcare

http://www.infobiomed.org/
INFOBIOMED

- Instrument: Network of Excellence
- Based on the success of the previous EC-funded BIOINFOMED study.
- Main objective: "Set a durable structure for Biomedical Informatics at the European level that supports its consolidation as an integrative scientific discipline that exploits the synergies between Biomedical Informatics and Medical Informatics."
- Scientific objectives:
  - progress in data interoperability, interfacing of methods, technologies and tools, pilot applications.
  - Integration of genetic and clinical information with preventive, diagnostic and therapeutic purposes.
ACGT
Advancing Clinico Genomic Trials on Cancer

http://www.eu-acgt.org
ACGT

- Duration: from 02/06 – to 01/10
- Instrument: Integrated Project
- Focuses on clinical trials on Cancer (Wilms tumor, Breast)
- Based on the principles of
  - Open access (among trusted partners)
  - Open source
- ACGT will deliver a European Biomedical GRID infrastructure offering seamless mediation services for sharing data and data-processing methods and tools, and advanced security.
- ACGT will trigger the emergence of latent *clinico-genomic synergies* to ensure *faster diagnosis* and more *efficient therapy*.
Objectives

• The ultimate objective of the ACGT project is the provision of a unified technological infrastructure which will facilitate:
  ▪ integrated access to multi-level biomedical data
  ▪ development or re-use of open source analytical tools, accompanied with the appropriate meta-data allowing their discovery and orchestration into complex workflows.
Data Access Infrastructure within ACGT
Data Access Architecture

The Action-Grid project is funded by the European Commission under the FP7
The ACGT Master Ontology on Cancer

- Priority: Ontology of entities involved in two carcinomas:
  - Nephroblastoma
  - Breast Carcinoma
- Terms from different terminologies and ontologies are integrated together
- Sources used as a dictionary
- Enhance cancer management in Europe by enabling semantic interoperability

The Action-Grid project is funded by the European Commission under the FP7
The ACGT Semantic Mediator

- Different components addressing different aspects of the same problem:
  - **Query Formulation Interface** → Helping end-users in formulating queries
  - **Master Ontology** → Acting as Global Schema
  - **Mediation Layer** → Resolving the query translation problem
  - **OntoQueryClean** → Dealing with query identifier heterogeneities
  - **OntoDataClean** → Addressing instance level heterogeneities
  - **Mapping API and GUI** → Aiding in the virtual views creation process
Scientific foundations of the Semantic Data Integration Approach

- **Query Translation** vs. **Data Warehouses**

- **Global as View** vs. **Local as View**
End user tools

- Clinical trial templates
  - Augmented by the ACGT Master Ontology
  - Graphical representation of data elements and processing steps in a clinical trial
- Data mining and analytical tools
- Clinical trial workflows
  - Graphical representation of workflows
  - Processing modules attached to each other in a workflow
ACTION-Grid
International Cooperative Action on Grid Computing and Biomedical Informatics

http://action-grid.eu
ACTION-Grid

• International Cooperative Action on Grid Computing and Biomedical Informatics between the European Union, Latin America, the Western Balkans and North Africa
• Duration: from 06/08 to 11/09
• Instrument: Support Action
Objectives

- **Main objective:** To establish a collaborative environment between the European Union, Latin America, the Western Balkans and North Africa in the Grid and Biomedical Informatics areas.

- **Scientific and technological subobjectives:**
  - To survey and analyse Grid-based and Biomedical Informatics initiatives in Europe, Latin America, the Western Balkans and North Africa.
  - To foster training and mobility in the involved regions in the areas of Grid computing, Biomedical Informatics and Nanomedicine.
  - To establish a future agenda covering the Grid/Nano/Bio/Medical Informatics.
  - To disseminate results through diverse means, e.g. conferences, articles, website, scientific journals, press releases, etc.
Survey and analysis

- Survey targeted to support the collaboration among different institutions in Europe, Latin America, the Western Balkans and North Africa in the areas of:
  - Research and development
  - Scientific and technological innovation
  - Training and mobility
  - Scientific dissemination
  - Technology transfer

- Build an index of Grid/Nano/Bio/ Medical resources
  - Huge number of resources publicly available over the Internet (databases, services and tools).
  - Problem of locating, identifying and selecting resources.
  - Compiles annotated resources and their associated information
  - Publicly accessible via web
Nanomedicine

Some areas and examples in Nanomedicine:

- Biopharmaceutics
- Implantable materials
- Implantable devices
- Nanosurgery
- Diagnostic tools
- Understanding basic life processes
- Modeling and simulation using informatics approaches
- Databases of nanoparticles.
Expected results & impacts

- To enhance synergy among Europe, Latin America, the Western Balkans and North Africa by facilitating the exchange of Grid-based methods and tools.
- To foster this synergies among all the actors in the ICT for Health area.
- To extend these synergies from Biomedical Informatics to the Nano-related areas.
Conclusions - Collaboration & Concertation

- The Biomedical Informatics Group has experience in:
  - Developing methods and tools for:
    - Knowledge discovery
    - Data mining and text mining
    - Management and integration in multiscale and heterogeneous biomedical data sources and biomedical data integration
    - Grid computing
  - Developing new project proposals and participate in Consortia for EC funded projects
- ACTION-Grid is planning to develop strong links with other projects for collaboration-harmonization, in areas such as:
  - Grid computing and middleware
  - Biomedical Informatics
  - Nanotechnology and Nanomedicine

The Action-Grid project is funded by the European Commission under the FP7
Acknowledgements

- European Commission
  - BIOINFOMED Study (EC-IST-2001-35024)
  - The INFOBIOMED NoE (FP6-2002-IST-507585)
  - The ACGT Integrated Project (FP6-2005-IST-026996)
  - The ACTION-Grid Support Action (FP7-ICT-224176)
Thank you!

Diana de la Iglesia
diglesia@infomed.dia.fi.upm.es
www.action-grid.eu