

# Easy and user-friendly workflow management based on the datamorphing concept

Stéphane GRAZIANI ISoft www.isoft.fr/bio



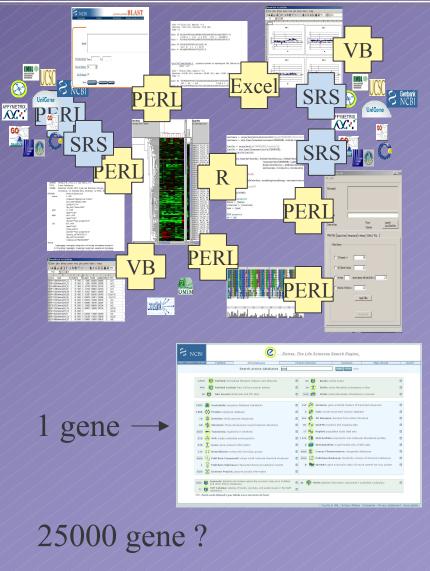
Amadéa Bio pack – © 2005



# Integration in Biology

# Data-morphing: an innovative approach for:

- Data integration
  - Very large volume
  - Heterogeneous data (Nature, format, access)
  - Concepts in constant evolution
- Tools integration
  - Heterogeneous formats
  - Different languages and environments
- Experimental methodology
  - From lab bench to web
- Evolutivity Reactivity
  - New data
  - New analyses





Amadéa Bio pack – © 2005



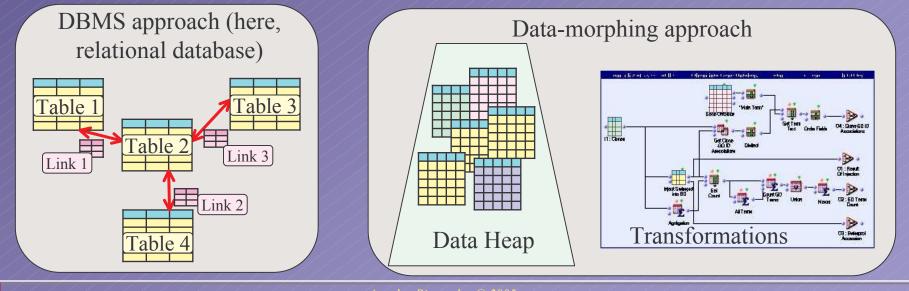
### Underlying concepts Fusion of Workflow and Dataflow

#### • Data-heap:

- Principle = <u>No a-priori structure</u> of the data
- Data is stored unstructured, reduced to the smallest atoms of information
- Data semantics is given by the way it is used

### Data-morphing

- Data-driven chaining of single transformation steps
- Very efficient data transformation engine (<u>1 million records per second</u>)
- <u>No limit</u> in the volume of treated data
- <u>No programmation</u>
- No need of a pre-existing data structure





Amadéa Bio pack – © 2005



# Easy pipeline definition

### A "biologist" speaking tool

- Definition of pipelines in an <u>homogeneous</u> environment
  - Focus on biological issues
  - Real-time ·
  - Gain in:
    - Quality
    - Productivity
    - Methodology
    - Execution time
  - Dramatically shorten experimentation cycles

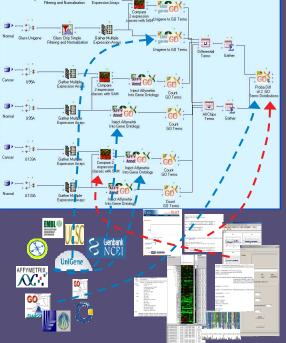
#### Features:

- <u>Instantaneous</u> connexion to any biological data sources
- <u>Intelligent linking</u> of any data source to any other
- Important panels of biological data analysis tools
- Library mechanism for easy non-programmatic extension

### Real-time

-Access to data -Data linking -Data analysis -Parameter tests -Hypotheses tests





# Data-morphing provides the <u>right data</u> at the <u>right place</u> at the <u>right moment</u>

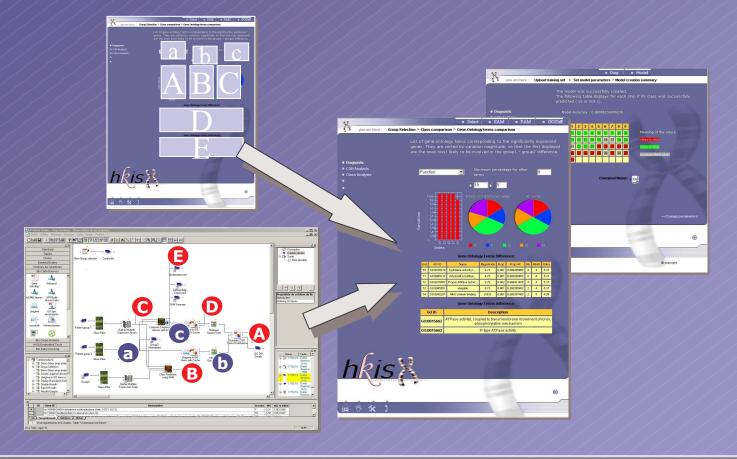


Amadéa Bio pack – © 2005



### Web interfaces

- Immediate publication of applications through the web
  - Interactive graphical interfaces for the end user





Amadea Bio pack – © 2003



- Quick overview of the platform
- How to develop an analysis pipe-line
  - Example : Starting from raw DNA chips data, normalize these data, and compare two classes of samples. From the differentially expressed genes, obtain:
    - A table showing per gene, with which disease it can be linked
    - The list of differentially distributed metabolic functions and the associated bibliography.
  - Steps:
    - Connexion to chips data
    - Normalization and construction of a summary table
    - Link to Refseq to obtain Gene information, including link to OMIM diseases
    - Link to OMIM to get disease title, and creation of a crossed table
    - Class comparison using SAM -> List of differentially expressed genes
    - Injection into the Gene Ontology Graph
    - Comparison with functions in the whole chip to get differentially distributed functions



**4madea**