

User-designed web services to support heterogeneous biological data retrieval

**Marie-Dominique DEVIGNES, Hervé de PALMA, Laurent PIERRON,
Lionel DOMENJOUR and Malika SMAIL-TABBONE**

*Laboratoire Lorrain de Recherche en Informatique et ses Applications (LORIA)
CNRS – INRIA - Université Henri Poincaré
Proliférateurs des Peroxysomes (EA3446) Université Henri Poincaré
Nancy, France*

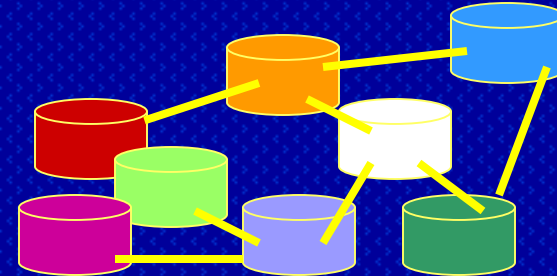


CONTEXT



One biological question

« Is this DNA fragment located upstream some interesting gene ? »



Hundreds of biological databases :
- interconnected but rarely interoperable
-- changing (content and presentation)

Querying

versus

Browsing

- Formulate a query
- Describe what you want

- ✦ Track the origin of the result
- ✦ Repeat the process

- ✦ Activate hyperlinks
- ✦ Discover what you cannot describe

- Disorientation
- Cognitive overload

Find a compromise between efficiency and exploration

OUTLINE

1. Heterogeneous biological data retrieval
 - A. User constraints
 - B. Case study
2. The Xcollect application
 - A. Principles
 - B. Architecture
3. Example - Demo
4. Encapsulation into applications

1. Heterogeneous biological data retrieval

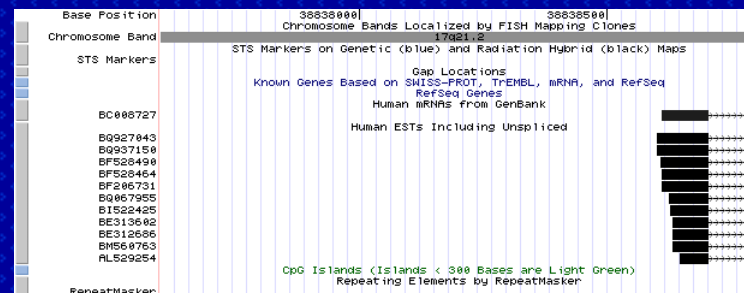
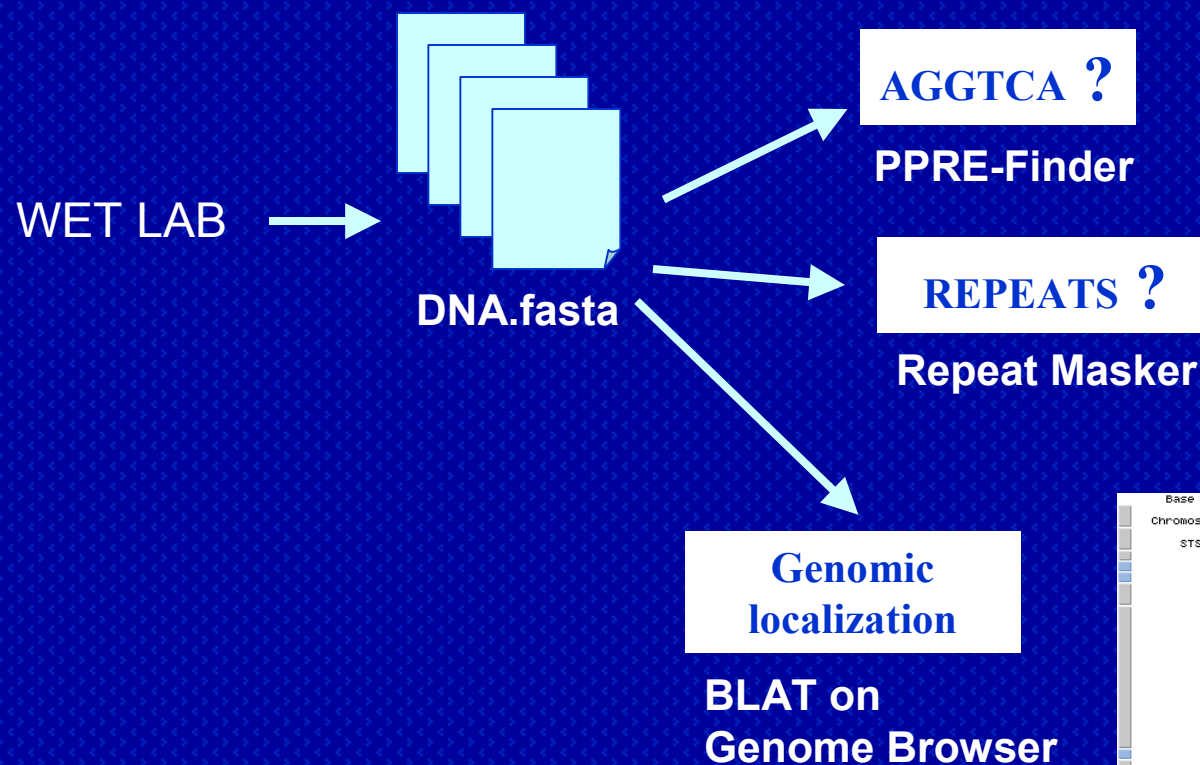
A . User constraints

- ✓ Unlimited choice of databases or treatments**
- ✓ Minimal maintenance costs**
- ✓ Easy adjustment to databases changes**
- ✓ Chaining of elementary steps (workflow)**
- ✓ Quality metadata retrieval**
- ✓ Generic solution**
- ✓ Automatic treatment of list of entries**
- ✓ Transformable session document**

B . Case studies (1/2)

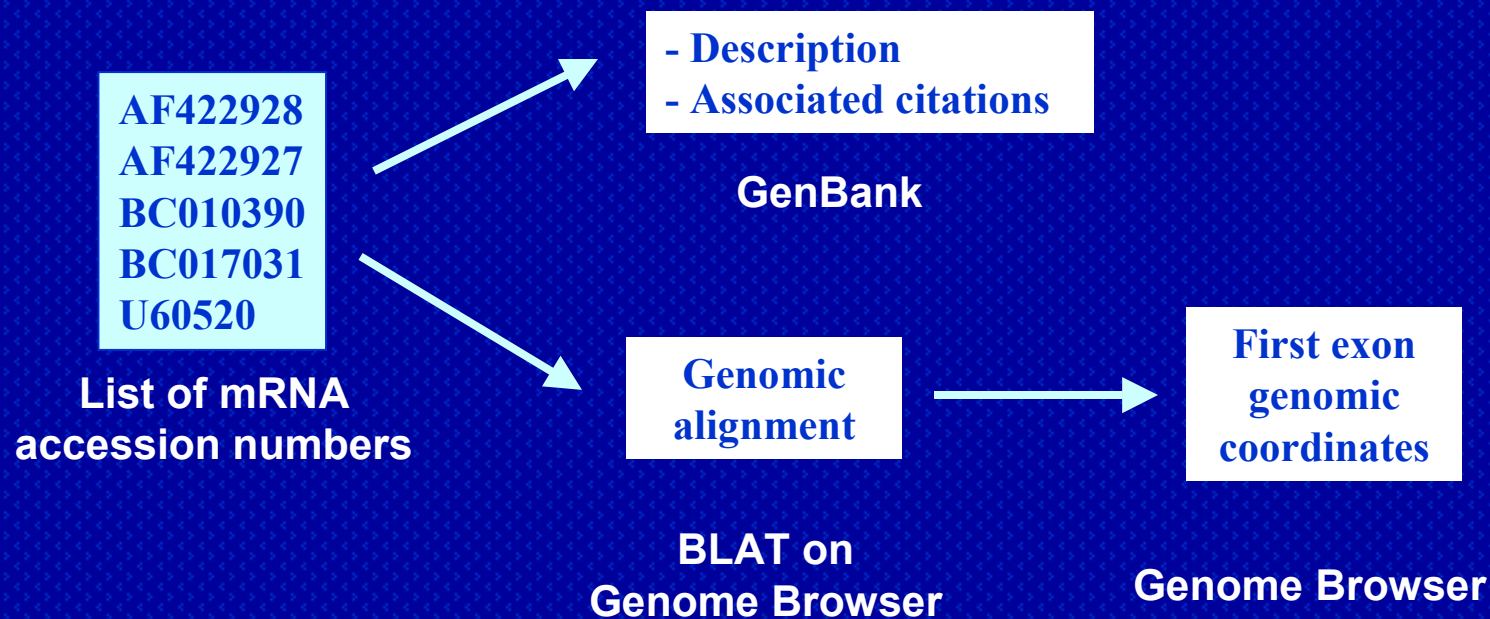
" Is this DNA fragment located upstream some interesting gene ? "

Xprom : From sequence to genomic context



B . Case studies (2/2)

XfirstExonMRNA-Hs



2. The « Xcollect » application

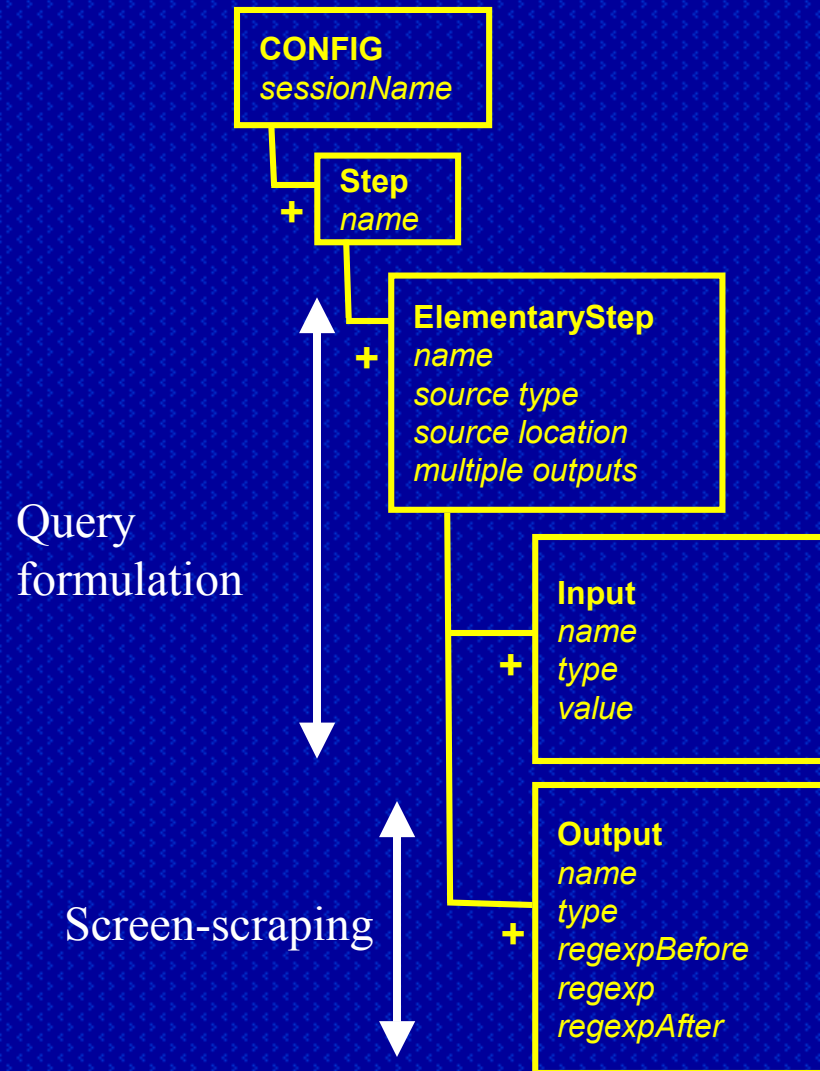
A. Principles :

1. Interaction with resources via HTTP protocol (URL queries, screen-scraping)
2. Distinguish scenario design and scenario enactment
2. Process any scenario (generic)

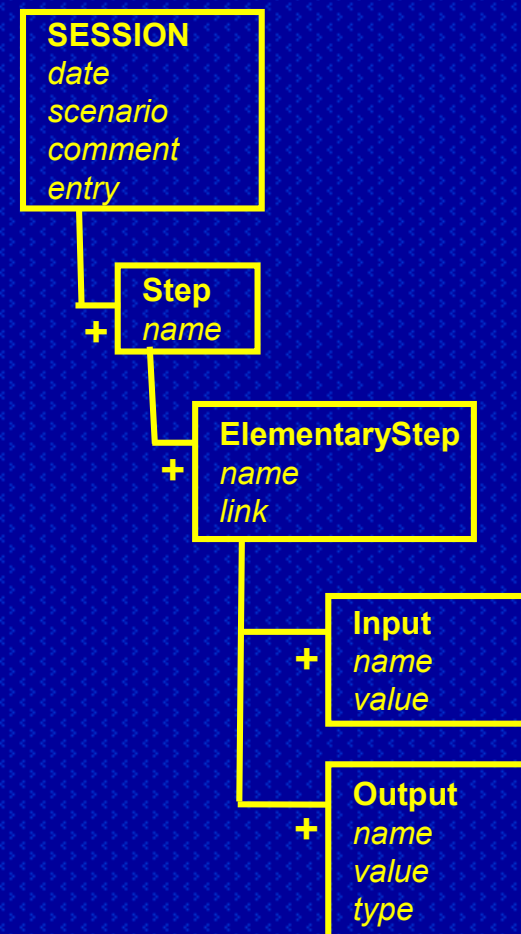
B. Models :

- ❖ XML scenario model
- ❖ Generic application that executes any scenario represented under this model
- ❖ XML session model
- ❖ Generic XSL sheet to display the results

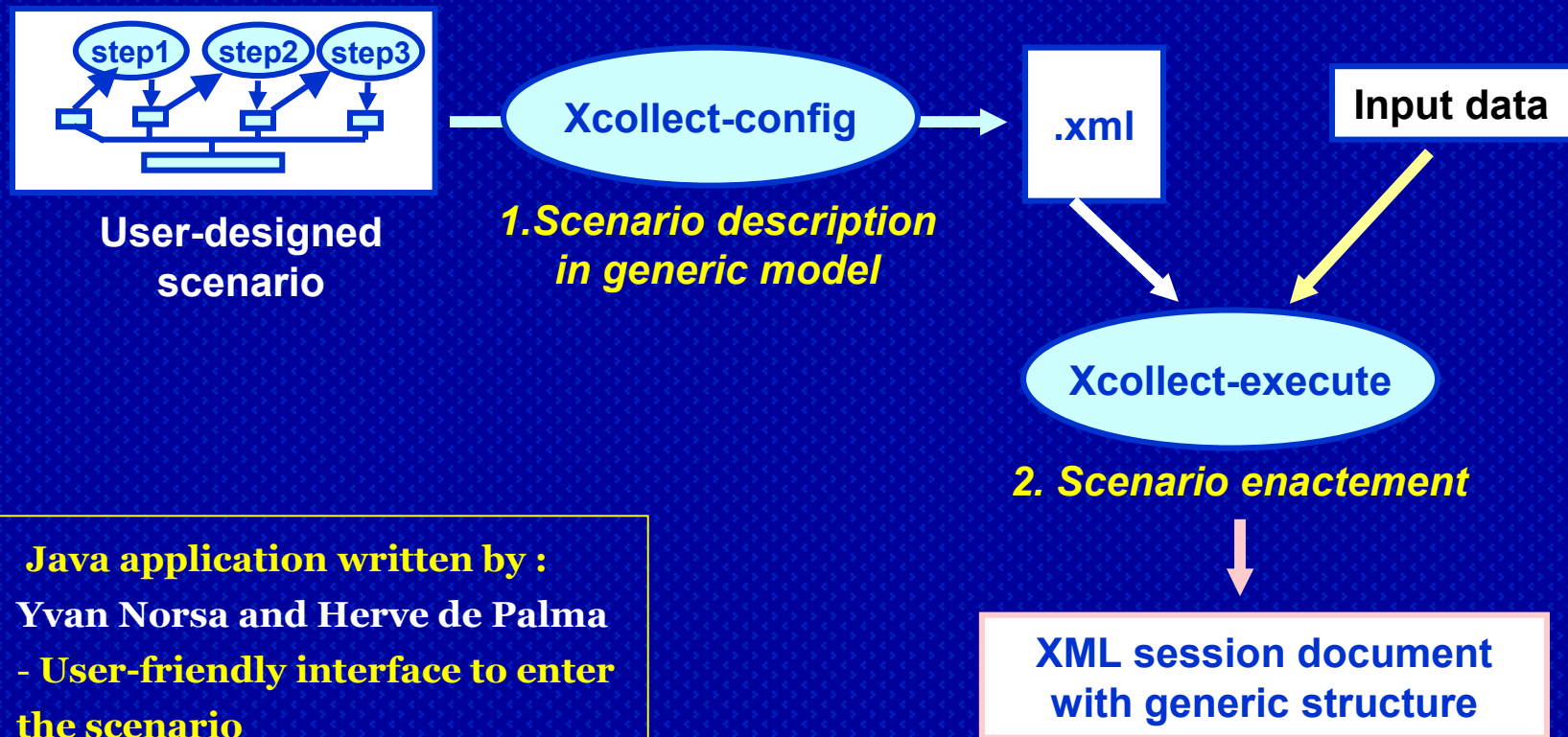
Generic scenario model



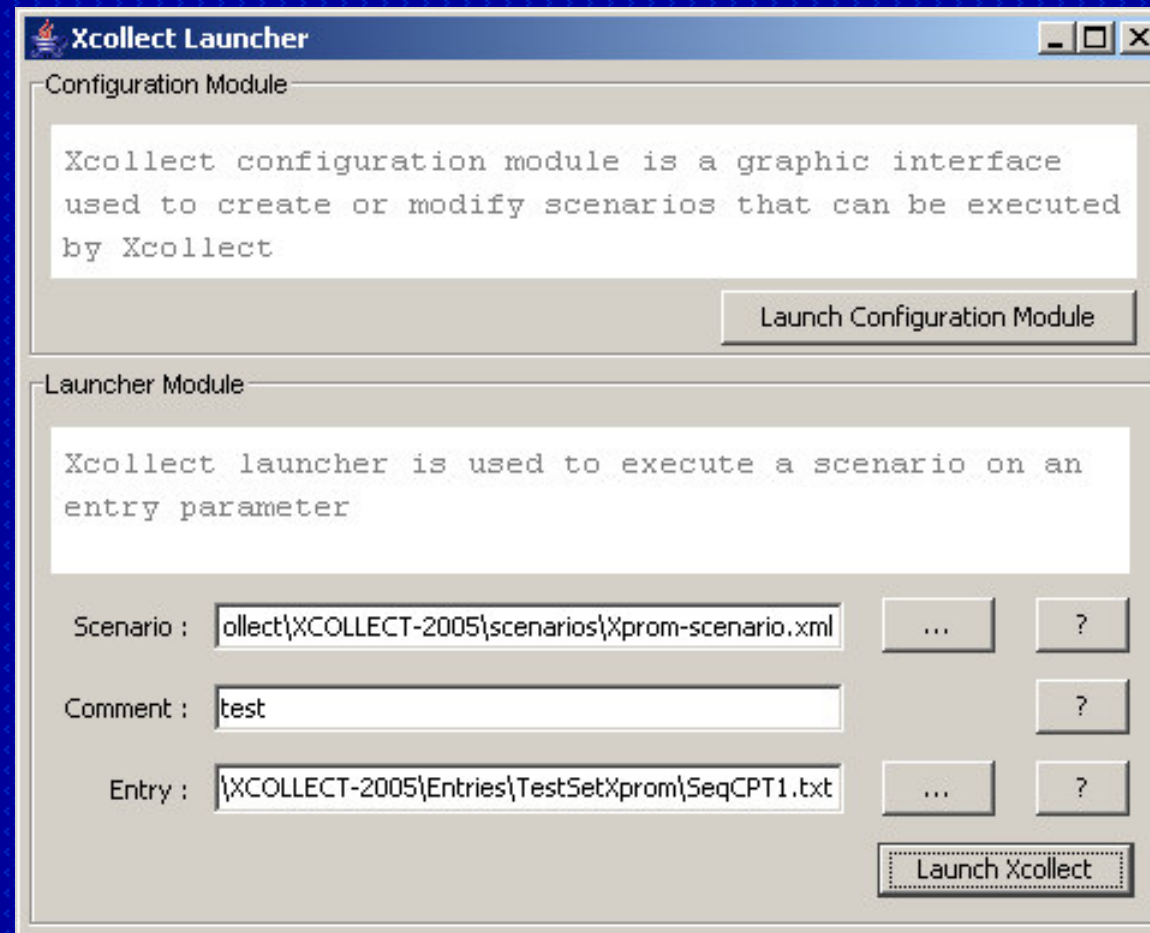
Generic session model



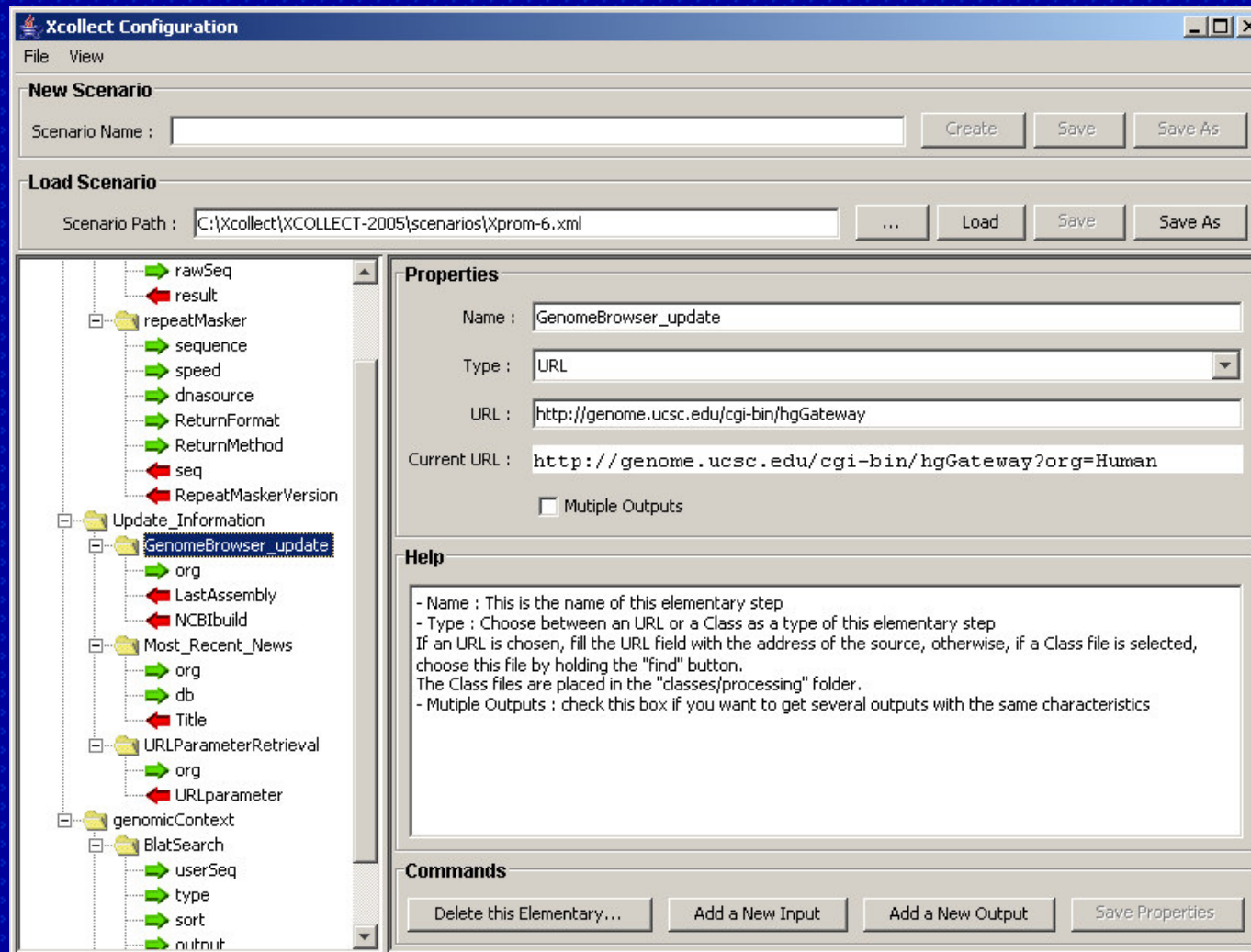
Schematic functioning



Xcollect interface



Configuration module



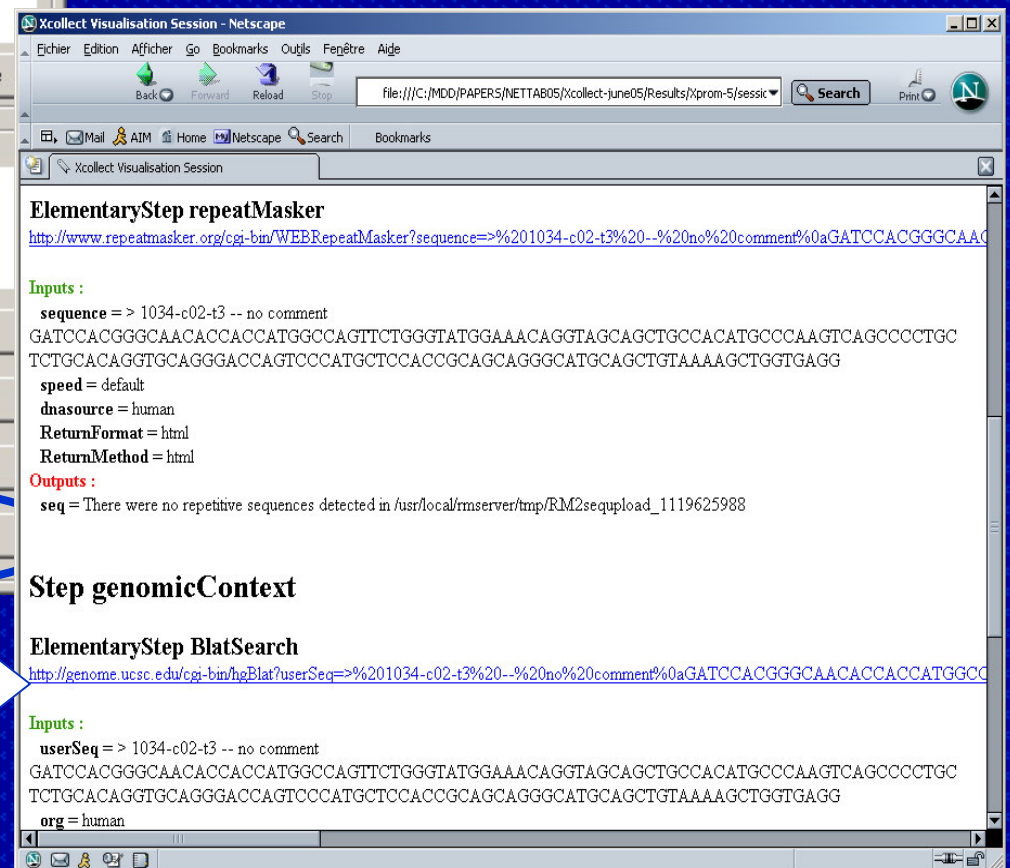
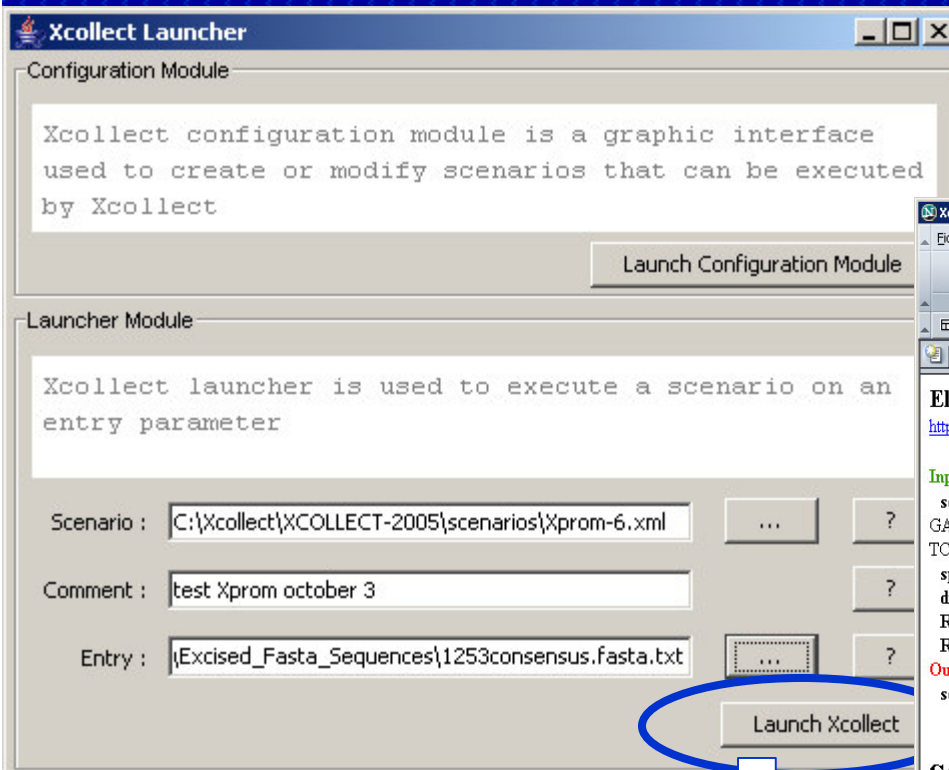
1. Heterogeneous biological data retrieval

2. Xcollect application

3. Example – Demo

4. Encapsulation into applications

Enactment module



2. Example – Demo

→ Go to DEMO

4. Encapsulation into applications

Web services

Goals :

- ✓ to facilitate re-use of scenarios
- ✓ to incorporate into higher-order workflows

Method

Xprom_ws.properties

display-name = getGenomic
Context
scenario = Xprom-scenario
output = XML file
input = sequence fasta file

description = Retrieves genomic
localisation from Genome Browser
and annotates by PpreFinder and
RepeatMasker

Xcollect_ws
deploy

Sun JWSDP
JAX-RPC
JAXR

ws
registry

Xcollect_ws
factory

Client : ws browser

Apache Tomcat

CONCLUSION - PERSPECTIVES

1. **Xcollect : a generic application to collect and integrate heterogeneous data from biological web sources :**

Along a « user-defined » scenario

Adaptable to any data source

2. **Make Xcollect scenarios available as web services**

3. **Semantic description of web services to facilitate their discovery**

Semantic « match-maker service » : e.g. Feta – myGrid

Analogous situation with biological databases when designing scenarios

→ BioRegistry project at LORIA and INIST (CNRS)

Thank you for your attention !