

Provenance Tracking and End-User Oriented Query Construction

Bartosz Baliś, Marian Bubak, Jakub Wach

Institute of Computer Science & ACC CYFRONET AGH, Krakow, Poland

<http://virolab.cyfronet.pl>

PROToS

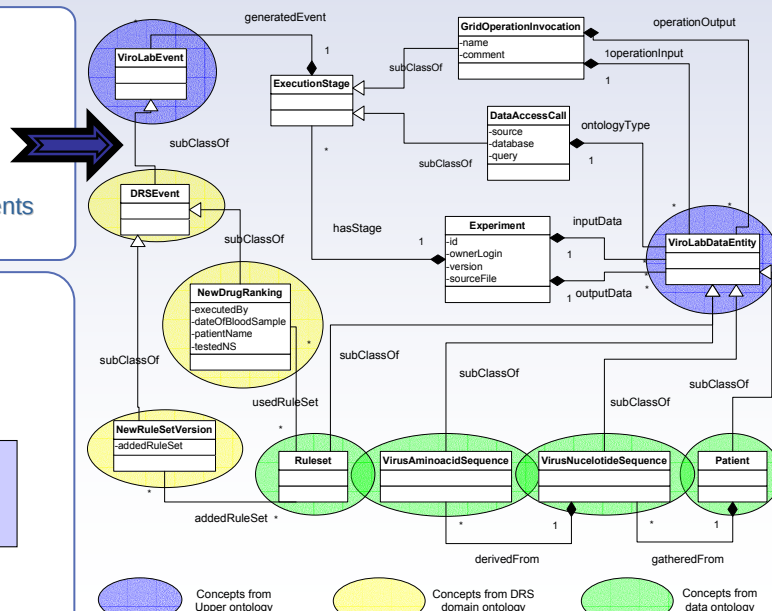
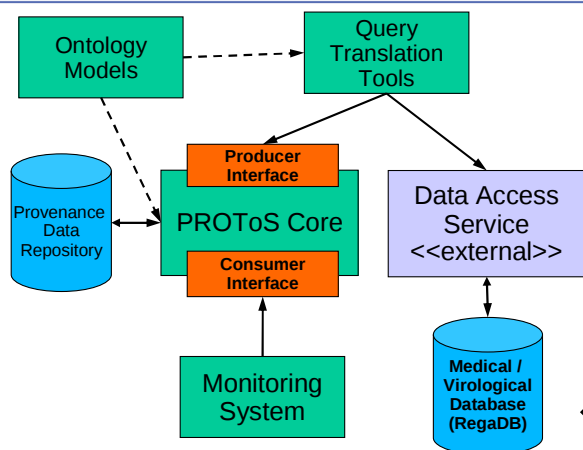
PROToS system for tracking, storing and querying provenance data of scientific experiments running in a ViroLab virtual laboratory

QUaTRO

QuATRO high-level ontology-based graphical tools to construct queries to data and provenance repositories in an end user-oriented manner

Ontology Models

- Experiment ontology represents specific execution of an experiment (core ontology)
- Domain ontologies enhance the semantic description
- Data ontology describes input / output data
- Service ontology describes individual services
- Application ontology describes application-specific events (e.g. DRS – Drug Ranking System application)



Architecture

- Events describing some aspect of an execution of an experiment are collected by a Monitoring System
- Low-level events are aggregated and published, through PROToS Core, as ontology individuals in a Repository
- QUaTRO Tools, using the Ontology Models, allow to construct queries to provenance and data repositories
- Supported scenario: Drug Ranking System – computation of best drug combination for a specific HIV virus sequence

Example QUaTRO Queries

- Begin with selecting a starting concept (e.g. Experiment)
- Tool dynamically loads ontology concepts, data model schemas and ontology individuals to expand the web form
- Provenance query: *select experiments which ...*
- Data query: *select virus aminoacid sequences which ...*

Query tree

Experiment >

☒ and ☐ and

hasStage

☐ and ☐ and

generatedEvent

☐ and ☐ and

executedBy

☒ and ☐ and

usedRuleSet

☐ and ☐ and

[select please...]
[select please...]
id
lastupdate
resourceid
aa_ii
version
location

Launch query

Query storage **Query results**

Query tree

VirusAminoacidSequence [select please...]

☐ and

[select please...]
[select please...]
derivedFrom
aa_sequence_ii
protein_ii
first_aa_pos
last_aa_pos

Launch query

Query storage **Query results**

Conclusion

- Ontology-based model of experiment information describes many aspects of experiment execution, including provenance
- Query tool enables non-IT specialists to construct complex queries over provenance and data

Conclusion

- **Ontology-based model of experiment information** describes many aspects of experiment execution, including provenance
- **Query tool enables non-IT specialists to construct complex queries over provenance and data**

virolab.cyfronet.pl/trac/protos

virolab.cyfronet.pl/trac/quatro

ViroLab EU-IST-027446

Coordinator: Prof. P.M.A. Sloot
Universiteit van Amsterdam
www.virolab.org

