

---

# A Platform for Collaborative e-Science Applications

Marian Bubak

ICS / Cyfronet AGH Krakow, PL

[bubak@agh.edu.pl](mailto:bubak@agh.edu.pl)

# Outline

---

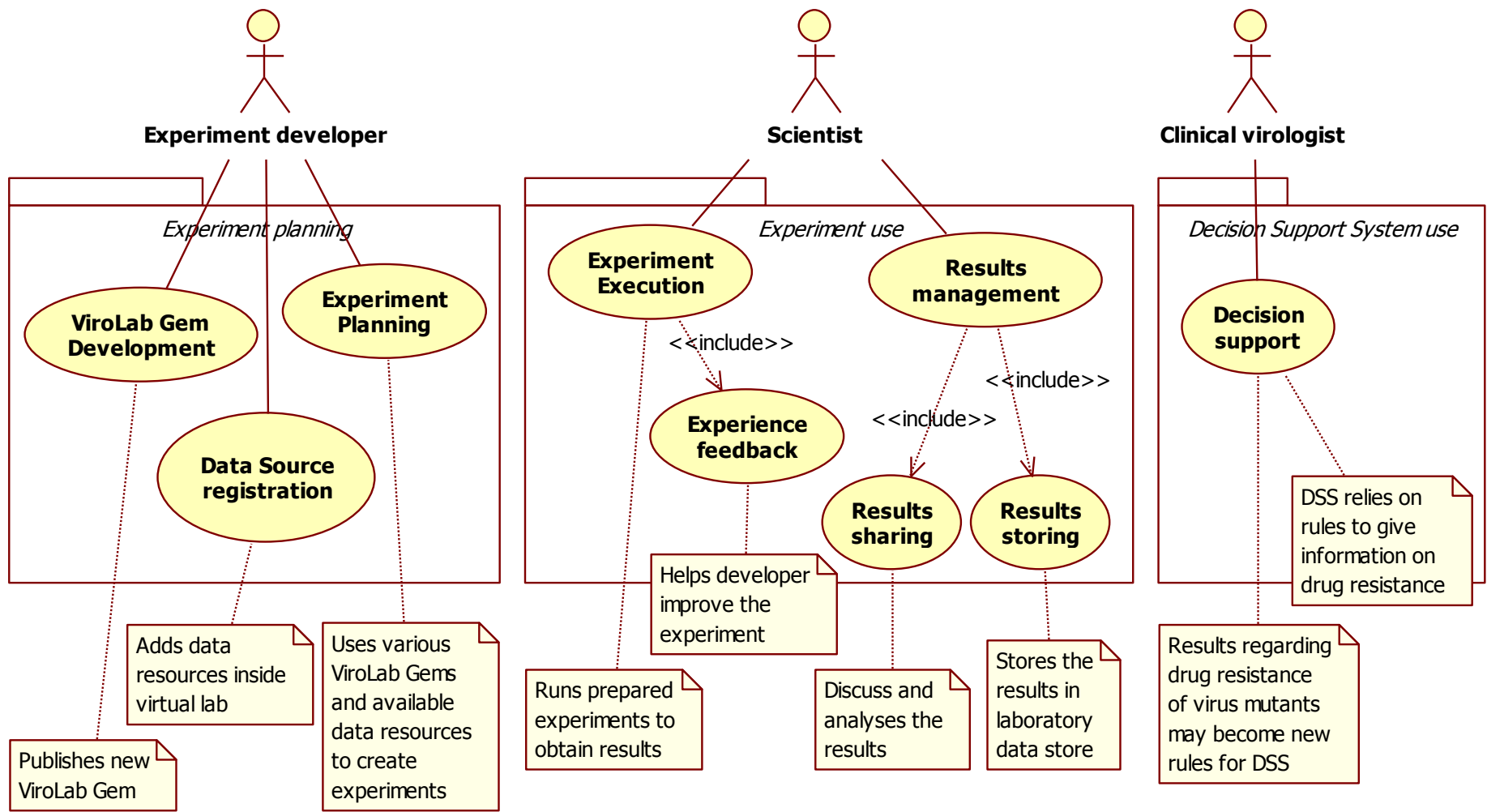
- Motivation
- Idea of an “experiment”
- Virtual laboratory
- Examples of experiments
- Summary and challenges

# System-Level Science

---

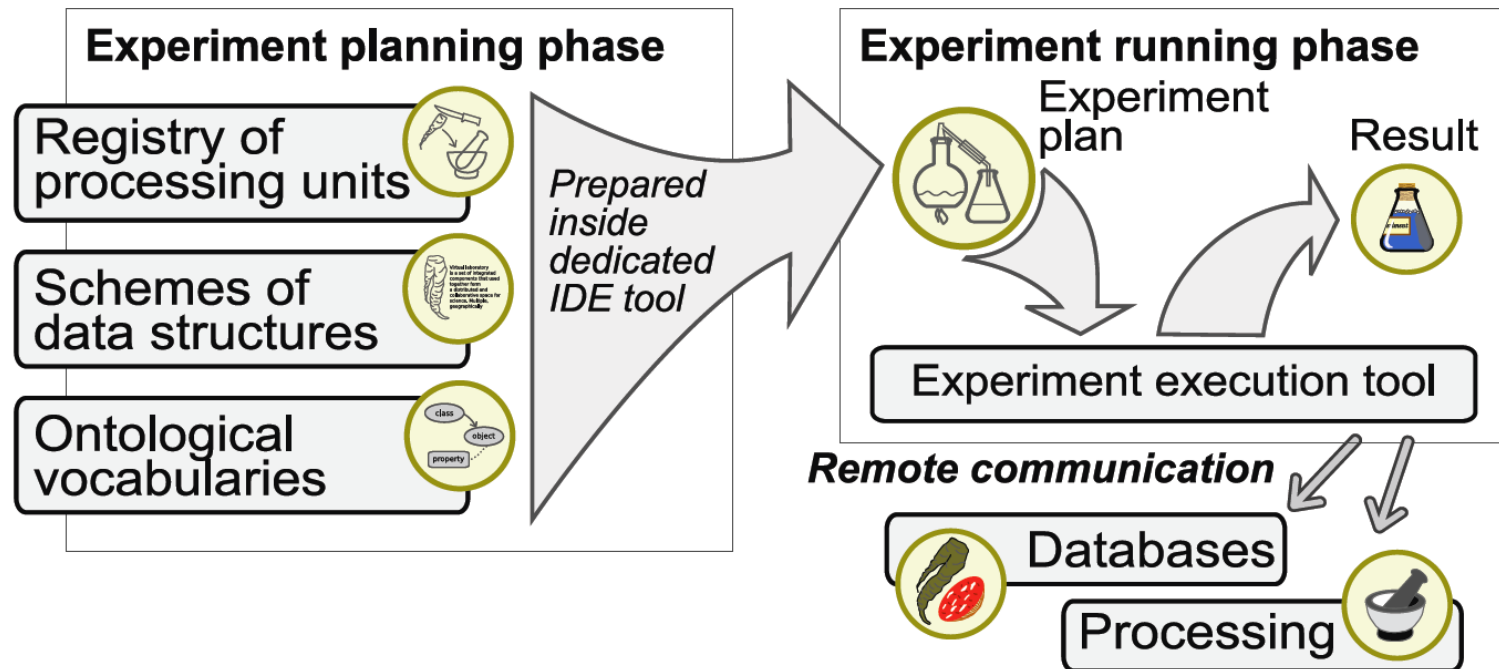
- An approach to scientific investigations which, besides of analyses of individual phenomena, integrates different, interdisciplinary sources of knowledge about a complex system, to acquire understanding of the system as a whole.
- Foster, I., Kesselman, C., *Scaling system-level science: Scientific exploration and its implications*. IEEE Computer 39 (11) 2006

# ViroLab Users



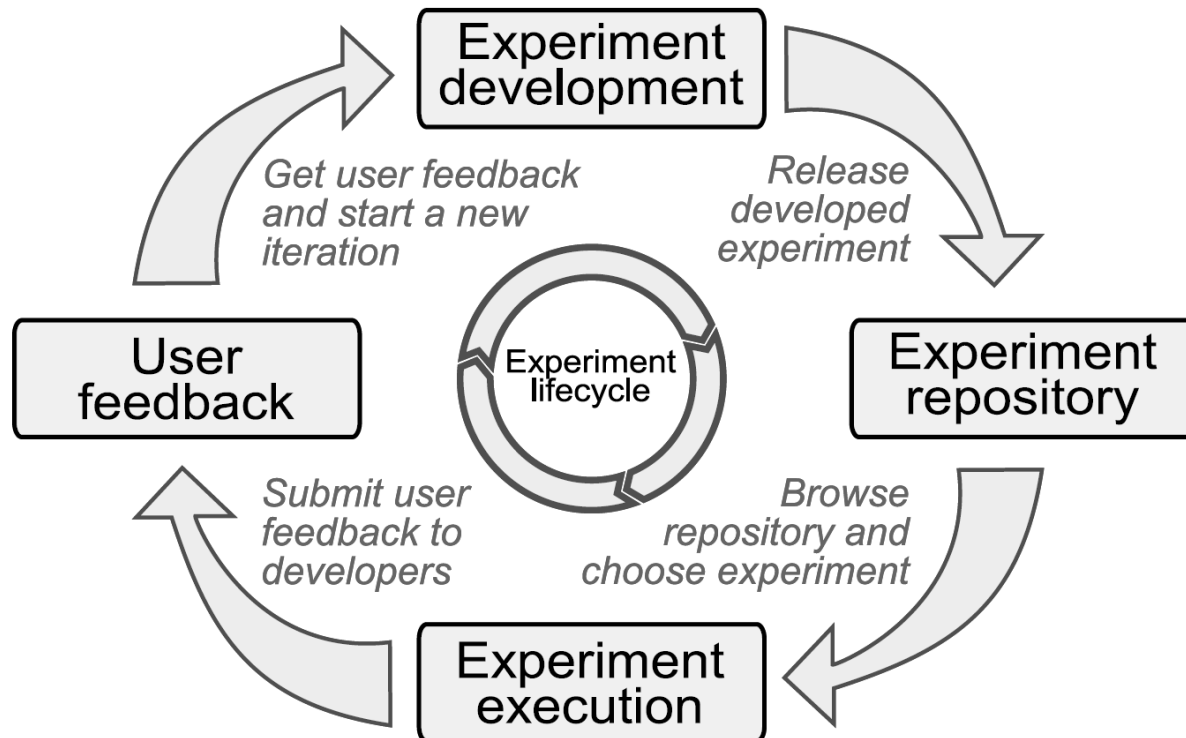
# Experiment

- **Experiment** - a process that combines together data with a set of activities that act on that data

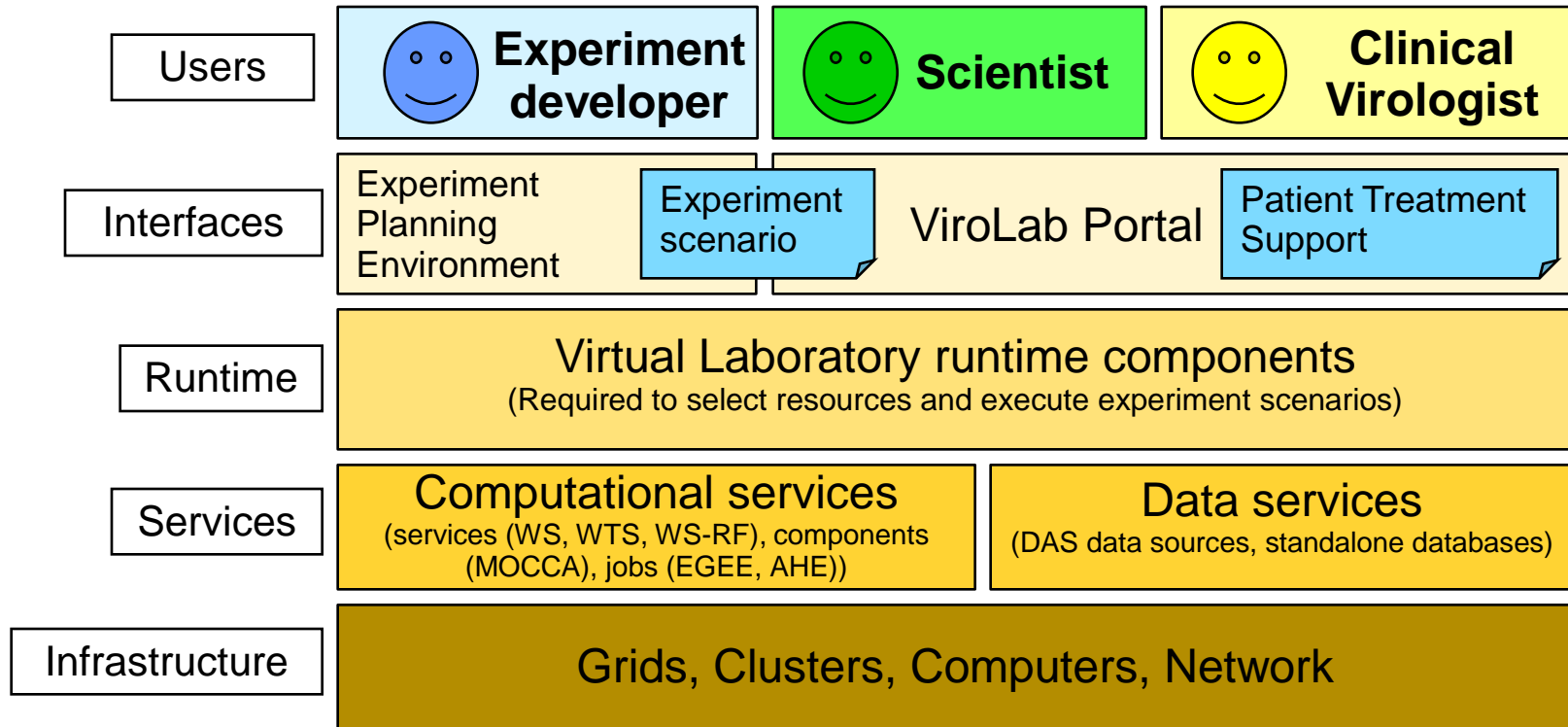


# Experiment Lifecycle

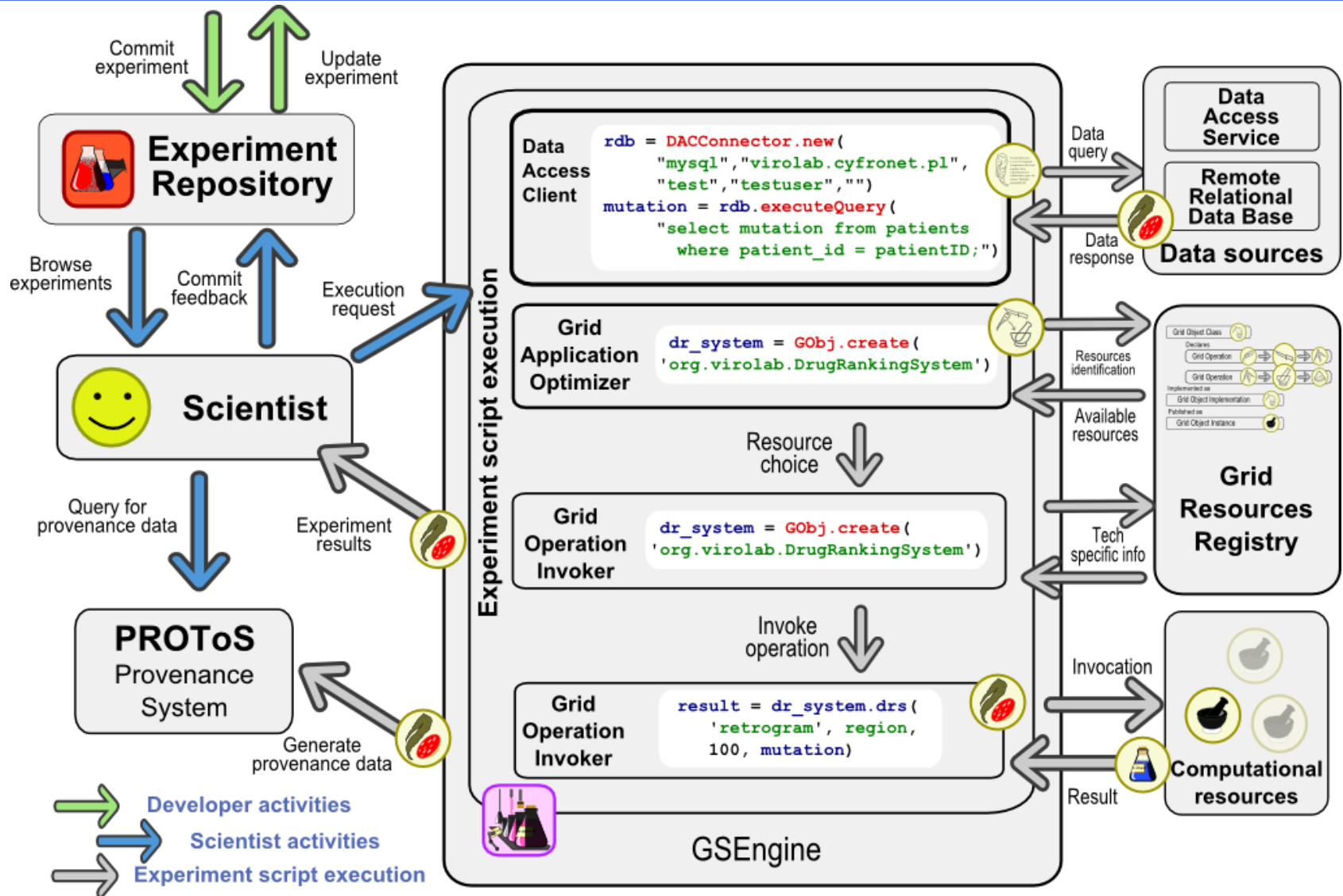
- **Experiment Pipeline** - is a collaborative planning and execution process that may create a new experiment



# On top of any infrastructure ...



# Collaborative Applications on VLvl

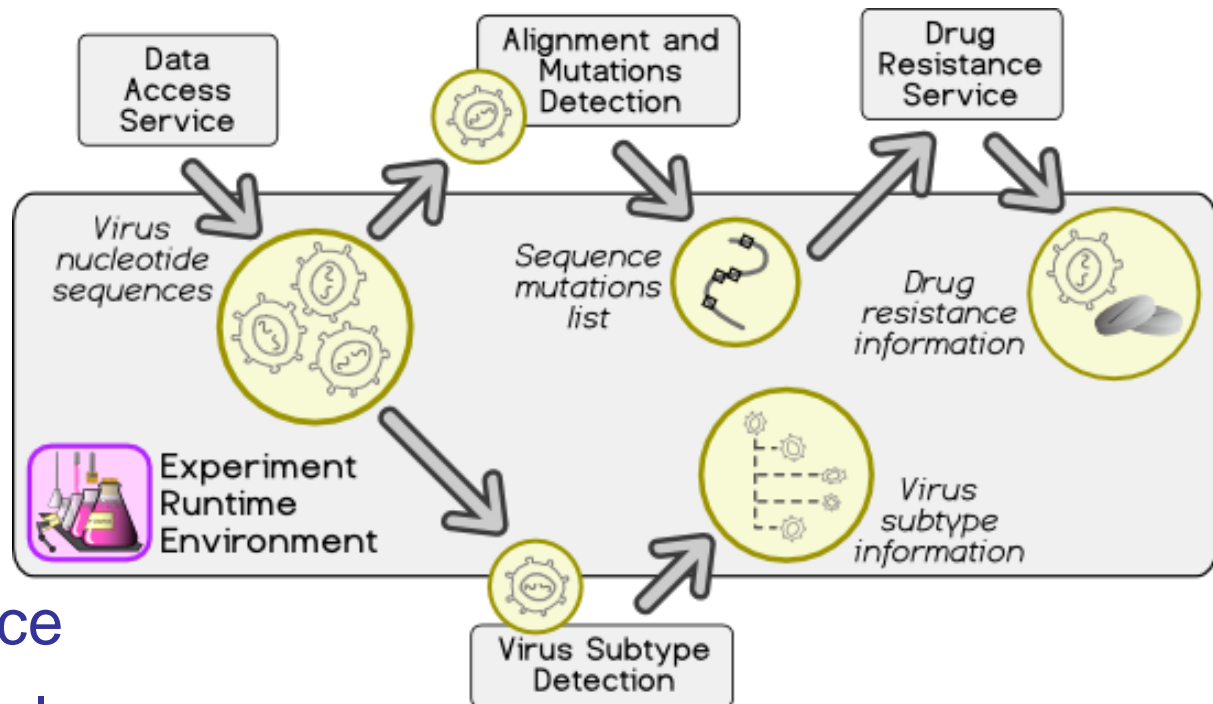




# Virus Genotype Analysis

**Objective:** loads nucleotide sequence of an HIV virus strain and provides its mutations and drug ranking information

- Gems used:
  - Alignment
  - Subtype detection
  - Drug ranking
  - Data Access Service
- Input: virus nucleotide sequence
- Output: various analyses



<http://virolab.cyfronet.pl/trac/vlvl/wiki/ExperimentDemo>

# Experiment Plan

patientID = 6 region = "rt"

**Parameters**

```
remoteDB = DACConnector.new( "DAS", "virolab.hlrs.de")
sequences = remoteDB.executeQuery(
  "select nucleotides from nt_sequence where
  patient_ii=#{patientID.to_s};")
```

**Genotype retrieval**

```
regaDBMutationsTool = GObj.create('regadb.RegadbMutationsTool')
regaDBMutationsTool.align(sequences, region.upcase)
mutations = regaDBMutationsTool.getResult
```

**Alignment + Mutation  
detection**

```
regaDBSubtypingTool = GObj.create('regadb.RegadbSubtypingTool')
regaDBSubtypingTool.subtype(sequences[0])
puts regaDBSubtypingTool.getResult
```

**Subtyping**

```
puts drs.drs('retrogram', region, 100, mutations)
```

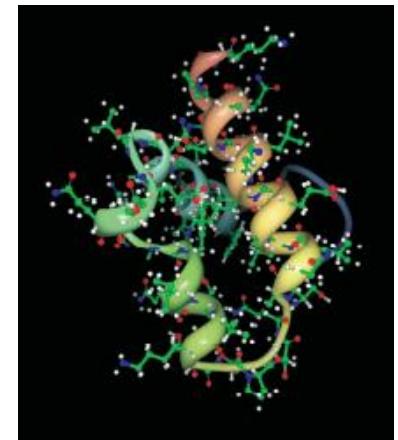
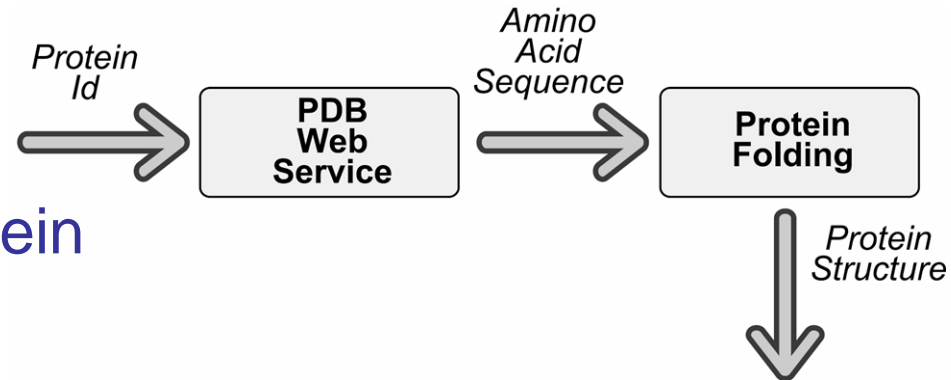
**Drug ranking**

# Protein Folding

**Objective:** demonstrate the usage of Virtual Laboratory for proteomics applications

- Input: protein and chain ID
- Output: 3D structure of protein
- Gems used:

- Protein Data Bank (PDB) Web Service
- Early-stage protein folding  
Bryliński M, Jurkowski W, Konieczny L, Roterman I.  
*Limited conformational space for early-stage protein folding simulation*  
Bioinformatics 20(2), 199-205 (2004)
- DAC and WebDAV for result storage

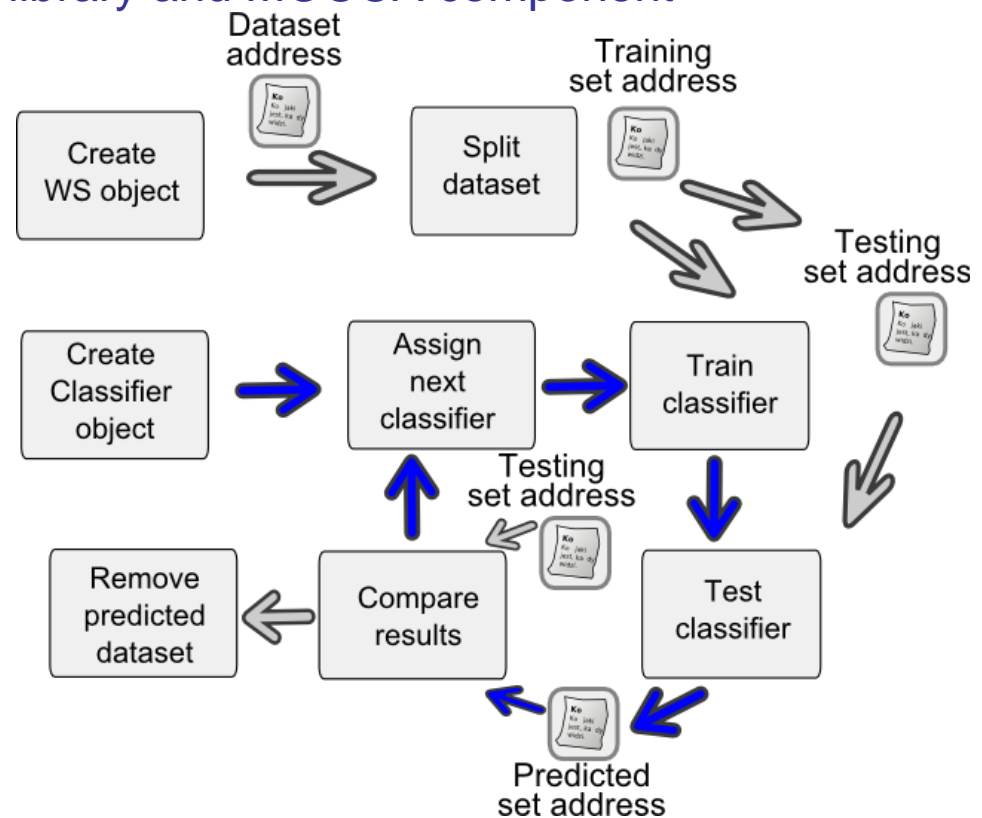


<http://virolab.cyfronet.pl/trac/exampleExperiments/wiki/exex/Folding>

# Data Mining with Weka

**Objective:** to analyze the quality of various classification algorithms on large datasets using Weka data mining library and MOCCA component framework.

- Input: sample dataset
- Output: quality of predictions
- Gems used:
  - Web services for data retrieval, conversion, splitting and testing
  - MOCCA components wrapping algorithms from Weka
  - WebDAV server for data storage



<http://virolab.cyfronet.pl/trac/exampleExperiments/wiki/exex/WekaAdv>

# Summary

---

- Collaborative applications
  - A new method of collaborative application development
  - Abstract layers to hide technological changes
  - Semantic description of applications
  - Integration of provenance recording and tracking
- Virtual laboratory
  - Semantic description of resources
  - Deployment on available Grid systems, clusters, and single CEs
  - Integration of WS, WSRF, components, jobs
  - Secure data access and integration

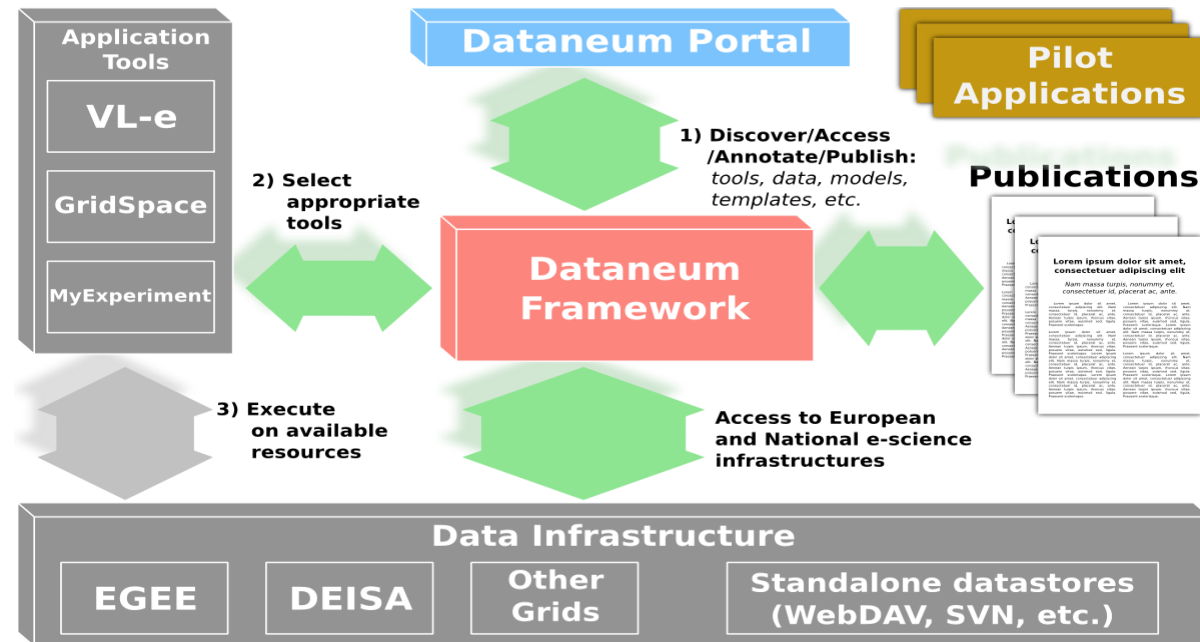
# Challenges

---

- development of Web-based tools for managing laboratories, running experiments, gathering results, refining methods and achieving scientific goals within virtual groups,
- to support collaborative aspects of research through dedicated, specialized tools for different groups of users (administrators, hardware providers, experiment developers, scientists) to complete their tasks within the overall goal of their groups,
- a collaborative space for scientific collaborative application development, facilitating publishing and sharing of software content.

# Dataneum

- **Authorization model** enabling e-Scientists to expose the input data and results of their research to the wider scientific community without violating the security restrictions inherent in grid computing infra-structures or collaborative working environments.
- **Integrated infrastructure and community web portal** for:
  - Authoring,
  - Publishing,
  - Managing,
  - Sharing,
  - Referencing,
  - Accessing,
  - Reusing,
  - Annotating, scientific data.



# Thanks to

---

- Runtime system
    - Tomasz Gubala, Marek Kasztelnik, Piotr Nowakowski, Eryk Ciepiela, Asia Kocot
  - Middleware
    - Maciej Malawski, Tomasz Bartynski, Jan Meizner
  - Presentation and collaboration tools
    - Włodzimierz Funika, Dariusz Krol, Daniel Harezlak, Alfredo Tirado
  - Data Access
    - Matthias Assel, Stefan Wesner, (Aenne Loehden, Bettina Krammer), Piotr Nowakowski
  - Provenance
    - Bartosz Balis, Jakub Wach, Michal Pelczar
  - Integration
    - Tomasz Gubala, Marek Kasztelnik
- 
- **VLvl description, demos, downloads** [virolab.cyfronet.pl](http://virolab.cyfronet.pl)
  - **ViroLab Project (Peter Sloot)** [www.virolab.org](http://www.virolab.org)