

# Architektura dużych projektów bioinformatycznych

Bartek Wilczyński

[bartek@mimuw.edu.pl](mailto:bartek@mimuw.edu.pl)

<http://www.mimuw.edu.pl/~bartek>

**Wykład 5. - Od systemów LI(M)S do Galaxy**  
9. XI. 2019

# Problem?



**"Think this is bad? You should see the inside of my head."**

# Zarządzanie danymi w bioinformatyce

- Dostępne i powtarzalne wyniki badań  
(Accessible and reproducible research)
- RREnvironment, RRSsystem, RRPublishing
- Systemy LI(M)S, Zarządzanie protokołami  
(Labkey, BASE, etc...)
- Literate programming (WEB, sweave, etc...)
- Systemy RRS – MeV, GenePattern,  
GenomeSpace, Galaxy

## COMPUTER SCIENCE

# Accessible Reproducible Research

Jill P. Mesirov

Scientific publications have at least two goals: (i) to announce a result and (ii) to convince readers that the result is correct. Mathematics papers are expected to contain a proof complete enough to allow knowledgeable readers to fill in any details. Papers in experimental science should describe the results and provide a clear enough protocol to allow successful repetition and extension.

Over the past ~35 years, computational science has posed challenges to this traditional paradigm—from the publication of the four-color theorem in mathematics (1), in which the proof was partially performed by a computer program, to results depending on computer simulation in chemistry, materials science, astrophysics, geophysics, and climate modeling. In these settings, the scientists are often sophisticated, skilled, and innovative programmers who develop large, robust software packages.

As use of computation in research grows, new tools are needed to expand recording, reporting, and reproduction of methods and data.



between two types of acute leukemia, based on language that can produce all of the text, figures,

# Pakiety LIMS

- Laboratory Information Management Software
- Zyskuje popularność od lat 80tych XX w. (era PC)
- Obecnie w zasadzie dwie możliwości:
  - desktop (w oparciu o lokalną bazę danych)
  - Klient-serwer, zwykle przez sieć www, choć istnieją jeszcze rozwiązania z lokalną bazą danych w modelu klient serwer

# Budowa pakietu LIMS

- Baza danych odczynników
- Baza przeprowadzonych analiz
- Linie komórkowe, populacje zwierząt laboratoryjnych
- Eksperymenty przeprowadzane na bieżąco
- Artykuły i rysunki, wyniki częściowe
- Coraz większe rozmiary danych pośrednich

# Pakiety LIMS

- Rynek rozczłonkowany pomiędzy bardzo wiele rozwiązań komercyjnych, często bardzo specjalizowanych:

Accelrys LIMS from Accelrys  
AgiLIMS from AgiLab  
ApolloLIMS from Common Cents Systems, Inc  
benchsys benchsys  
Biotracker from Ocimum Bio Solutions  
Biotracker Lite from Ocimum Bio Solutions  
CaliberLIMS from Caliber Technologies Pvt. Ltd.  
Care Med LIS from Medcare International  
CCLAS from Ventyx, an ABB company, formerly Mincom (company)  
Clarity LIMS from GenoLogics Life Sciences  
CloudLIMS from CloudLIMS  
Cyberlab in Cloud (Toplab) from Megaweb websites: Megaweb and TOPLAB  
Darwin from Thermo Fisher Scientific  
ELab from LabLynx  
Element LIMS from Promium  
eQMS::LIMS Pardus d.o.o. [1]  
Exemplar Biomarker Discovery from Sapio Sciences  
Exemplar Dx LIMS from Sapio Sciences  
Exemplar Research LIMS from Sapio Sciences  
Galileo from Thermo Fisher Scientific  
LABAsistan from Tenay Medical Software

LABbase from Analytik Jena  
LabPlus PRÉVENTION EXPERT CONSEIL INC. (PEC)  
LabSoft LIMS from Computing Solutions Inc.  
LABVANTAGE from LABVANTAGE Solutions  
Labware from LabWare  
LABWORKS from PerkinElmer  
Labway-LIMS from Ambidata Digital Innovation Solutions & Consulting  
LDMS from Frontier Science and Technology Research Foundation  
Matrix Gemini from Autoscribe  
MetaField Lab from Agile Frameworks, LLC  
Nautilus from Thermo Fisher Scientific  
ProlabQ from Open-Co  
readyLIMS from Analytik Jena  
Result Point from Accelerated Technology Laboratories, Inc  
SampleManager from Thermo Fisher Scientific  
Sample Master from Accelerated Technology Laboratories, Inc  
Select Agent Inventory (SAI) Management System Foxspire  
SchuyLab from Schuyler House, Inc  
SIMATIC IT Unilab from Siemens  
SLims from Genohm  
Solution Laboratoire from Limseo  
SmartLims from SmartSoft, Inc  
STARLIMS from STARLIMS Corporation  
StrainControl Laboratory Manager from DNA Globe  
TITAN from Accelerated Technology Laboratories, Inc  
TremoLAB from Binsol S.A. [www.binsol.com.ar](http://www.binsol.com.ar)  
Watson from Thermo Fisher Scientific  
webLIMS from LabLynx  
WinLIMS from QSI Corporation N  
NuGenesis from Waters Corporation [www.Waters.com](http://www.Waters.com)

# Od niedawna również oprogramowanie open source

- Labkey Server (apache license)
- MISO (GPLv3)
- BIKA LIMS (AGPLv3)
- Typowy model to komercyjna firma rozwijająca oprogramowanie i świadcząca usługi wsparcia
- Często dużo tańsze od rozwiązań komercyjnych, ale wymagające większego know-how na miejscu, popularne w dużych instytucjach



# Pakiety typu open lab notebook

... there is a URL to a laboratory notebook that is freely available and indexed on common search engines. It does not necessarily have to look like a paper notebook but it is essential that all of the information available to the researchers to make their conclusions is equally available to the rest of the world

- —Jean-Claude Bradley



# Dla eksperymentów mikromacierzowych

- System BASE

**BMC Bioinformatics**



Software

**Open Access**

## **BASE - 2nd generation software for microarray data management and analysis**

Johan Vallon-Christersson<sup>1,2</sup>, Nicklas Nordborg<sup>3</sup>, Martin Svensson<sup>3</sup> and Jari Häkkinen\*<sup>1</sup>

Address: <sup>1</sup>Department of Oncology, Clinical Sciences, Lund University, SE-221 84 Lund, Sweden, <sup>2</sup>CREATE Health Strategic Centre for Translational Cancer Research, Lund University, SE-221 84 Lund, Sweden and <sup>3</sup>Department of Theoretical Physics, Lund University, Sölvegatan 14a, SE-223 62 Lund, Sweden

Email: Johan Vallon-Christersson - johan.vallon-christersson@med.lu.se; Nicklas Nordborg - nicklas@thep.lu.se; Martin Svensson - martin@thep.lu.se; Jari Häkkinen\* - jari.hakkinen@med.lu.se

\* Corresponding author

Published: 12 October 2009

BMC Bioinformatics 2009, 10:330 doi:10.1186/1471-2105-10-330

Received: 23 June 2009

Accepted: 12 October 2009

# Do eksperymentów sekwencjonowania

- Galaxy server

The screenshot displays the Galaxy web interface. At the top, there's a navigation bar with options like 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Cloud', 'Help', and 'User'. Below this is a search bar for tools. The main content area features a central banner that reads 'Try Galaxy on the Cloud' with a subtext 'Now you can have a personal Galaxy within the infinite Universe'. To the right of the banner is a 'Tweets' section showing three tweets from the Galaxy Project (@galaxyproject). On the far right, there's a 'History' sidebar showing 'Unnamed history' with 0 bytes and a message that the history has been deleted. At the bottom of the page, there are logos for Penn State, Johns Hopkins University, TACC, and iPlant Collaborative, along with text describing the Galaxy Team's affiliation and the infrastructure provided by TACC and iPlant Collaborative.

**Galaxy** is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy [start here](#) or consult our [help resources](#).

**Try Galaxy on the Cloud**

Now you can have a personal Galaxy within the infinite Universe

**Tweets**

- Galaxy Project** @galaxyproject 41m  
Fall 2014 Galaxy User Group Grand Ouest (GUGGO) Events Report: Tools, User Group, RAD-Seq: [bit.ly/1vquYOm](http://bit.ly/1vquYOm) #usegalaxy @Biogenouest
- Galaxy Project** @galaxyproject 31 Oct  
Next Generation Data Analysis Workshop, Dec 5-8, 2014 @UCRiverside [bit.ly/ucrworkshops](http://bit.ly/ucrworkshops) #usegalaxy  
Expand
- Galaxy Project** @galaxyproject 31 Oct  
Research Specialist, Institute for Cyber-Enabled Research, Michigan State University, United States [bit.ly/13pqd0T](http://bit.ly/13pqd0T) #usegalaxy  
Expand

Tweet to @galaxyproject

**History**

search datasets

**Unnamed history**  
0 bytes

This history has been deleted

This history is empty. You can [load your own data](#) or [get data from an external source](#)

**PENNSYLVANIA STATE UNIVERSITY**

**JOHNS HOPKINS UNIVERSITY**

**TACC**

**iPlant Collaborative**

The Galaxy Team is a part of the [Center for Comparative Genomics and Bioinformatics](#) at Penn State, and the [Department of Biology](#) and at [Johns Hopkins University](#).

This instance of Galaxy is utilizing infrastructure generously provided by the [iPlant Collaborative](#) at the [Texas Advanced Computing Center](#), with support from the [National Science Foundation](#).