New approach to Genomics Experiments
Taking Advantage of Virtual Laboratory System

Luiza Handschuh[2, 3], Marcin Lawenda[1], Piotr Stępniak[2],
Marek Figlerowicz[2], Maciej Stroiński[1] and Jan Węglarz[1]

[1] *Poznań Supercomputing and Networking Center*
Noskowskiego 10, 61-704 Poznań, Poland

[2] *Institute of Bioorganic Chemistry PAS*
Noskowskiego 12/14, 61-704 Poznań, Poland

[3] *Poznań University of Medical Sciences, Departament of Hematology*
Szamarzewskiego 84, 60-569 Poznań, Poland
e-mail: vlab@man.poznan.pl

Abstract: Specialized software, on-line tools and computational resources are very common in contemporary science. One of the exemplary domain is genomics – a new branch of science that developed rapidly in the last decade. As the genome research is very complex, it must be supported by professional informatics. In a microarray field the following steps cannot be performed without computational work: design of probes, quantitative analysis of hybridization results, post-processing, and finally data storage and management. Here, the general aspects of virtual laboratory systems are presented together with perspectives of their implementation in genomics in order to automate and facilitate this area of research.

Key words: genomics, virtual laboratory, remote instrumentation

I. INTRODUCTION

There are numerous areas of science, industry and commerce that require broad international co-operation for their success. A number of problems may be addressed by using sophisticated equipment and top-level expertise, which is often locally unavailable. Therefore, the development and dissemination of techniques and technologies that allow virtualized, remote and shared access to industrial or scientific instruments is essential for the progress of society. The possibility of using scientific or industrial equipment independently of their physical location helps in the egalitarian attitude in using expensive facilities and for unification of communities, and subsequently opens new opportunities for industry, science and business. That is why a new branch of grid science – Remote Instrumentation Systems (RIS) – has become so important for balanced development of expensive equipment-required domains.

II. VIRTUAL LABORATORY

The specific representatives of the RIS systems are virtual laboratory systems (VL). One of them is PSNC Virtual Laboratory (VLab) [2], a research project developed in Poznań Supercomputing and Networking Center [9] since the beginning of the year 2002.

In general, VL is a distributed environment providing remote access to various kinds of scientific equipment and computational resources. In VLab users can submit their tasks in a form of a Dynamic Measurement Scenario –sets of experiments and computational tasks of different kind. These tasks form a dependencies graph describing the connections between them and all possible flow paths; the actual flow path is determined upon run-time based on results obtained on each stage. The tasks are scheduled on a one-by-one (or group) basis, after the appropriate results from the preceding tasks are obtained. The Virtual Laboratory is not a standalone system. It was designed to cooperate with many other grid systems, providing only the purpose-specific functionality and relying on well-known and reliable grid solutions. The most important system the VLab cooperates with is the Globus Toolkit [16] – in the scope of scheduling computational tasks, software services and libraries for resource monitoring, discovery, and management. All computational tasks submitted in the VLab system are transferred to the Globus via the GRMS [17] module – an important part of the GridLab project [18].

Among other external systems used by the Virtual Laboratory are: the VNC [19] system, SZD [1] (data management system), authentication and authorization modules.

II.1. Architecture

The Virtual Laboratory system has a modular architecture. The modules can be grouped into three main layers. The general diagram is presented in Fig. 1.
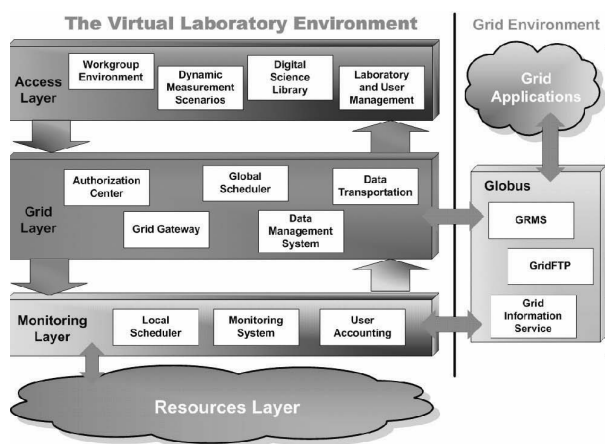

Fig. 1. General architecture of the Virtual Laboratory

An Access Layer is the top layer of this structure. Basically, it contains all modules and components responsible for a VLab user access and graphical interface to the system (including a web portal), and data input interface. Below there is a Grid Layer which communicates with the external grid environment. It is responsible for user authorization and authentication, data management, and general task scheduling. Modules from this layer also handle the transfer of the computational tasks to the Globus system, and gather feedback data and the computed results. The Monitoring Layer consists of lower-level modules, such as hardware dedicated schedulers, system monitoring, gathering accounting data, etc. All the physical scientific devices are located in the Resources Layer, as well as modules responsible for their direct control.

On the Grid Environment side the most important element is the Globus system with all its components (GRMS, GridFTP etc). Globus also allows to execute computational tasks (batch and interactive) on a wide variety of grid applications.

II.2. The Peculiar Nature of the Virtual Laboratory Experiments

To explain how the VLab works, the peculiar nature of this system must be described first.

In general, there are two main kinds of tasks in the Virtual Laboratory: experimental and computational ones. The latter can be divided into regular (batch) jobs and interactive/visualization tasks (the ones performed in the real time, directly by the users via the GUI). The biggest difference between those types is that in the interactive tasks the time slot reserved for running the task on a computational machine must be synchronized with user preferences, considering specific work hours, daily schedule, etc. Another aspect is the mechanism which will present the users with the graphical interface of the actual computational (or visualization) application – which is run on dynamically assigned computational server – and allow them to perform their interactive task.

Another very specific type of Virtual Laboratory tasks are the experiments. By the term *experiment* we mean a task scheduled to be performed on the remote laboratory equipment, available via VLab to its users. In most cases such experiments will be interactive processes, with users manipulating directly the remote equipment via specialized control software GUI. In every science domain there can be many dependencies of external, often indeterministic factors. The device will not be available for VLab users non-stop but will be shared with local researchers (usually with higher priority than the remote users). There are also maintenance periods in which the device is unavailable. Sometimes the presence and assistance of the device operator may be necessary – especially at the beginning of experiments.

The first scientific device incorporated into the VLab system was an NMR spectrometer. The most important problems with scheduling the NMR experiments, apart from those described above, come from the samples management. To perform an NMR experiment, an actual sample containing a chemical compound has to be delivered to the NMR spectrometer and inserted into the machine. At the task (and the corresponding sample) submission point the actual NMR device has to be known and chosen, because the sample has to be sent from the remote location to the device site. The exact time of a sample arrival is not known as well, making the exact task scheduling impossible until the sample arrives.

II.3. Workflow Management

Experiments executed in the science laboratories are complex and consist of many stages. Usually it looks as follows: a scientist prepares a sample and/or input data (e.g. parameters) which will be measured/computed. Next s/he uses laboratory devices to obtain data which are then processed by specialized software. Processing can include

the visualization stage if it is needed to assess the results. In case of undesirable results some measurement stages should be repeated.

At each stage the scientist decides which way the research should go next. As we can see, the experiment process execution may consist of very similar stages in many scientific disciplines (laboratories). The given experimental process is often executed recurrently by some parameters modification. Obviously, the presented scenario is typical but we realize that scholars can have more sophisticated needs.

Thus, we can create a graph which describes the execution path specified by the user. Nodes in this graph correspond to experimental or computational tasks. Edges (links) correspond to the path the measurement execution is following. In the nodes we have a defined type of application and its parameters. In links the passing conditions which substitute decisions made by the user are defined. These conditions are connected with applications, and let us determine if the application is finished with the desired results (if not, the condition is not met). This execution graph with specified nodes and links (with conditions) is called the Dynamic Measurement Scenario (DMS). The term "dynamic" is used, because the real execution path is determined by the results achieved on each measurement stage and can change.

To properly define the Dynamic Measurement Scenario, the system has to have knowledge about available applications and connections which are enabled to create. In case of the laboratory type, where the processing software is well known, it seems to be easy (an example shown in Fig. 2). The issue will be more complex when we want to define the DMS model for a wider range of virtual laboratories.

To solve this problem we have defined a special language – Dynamic Measurement Scenario Language (DMSL) which determines the following parameters: names of the connected applications, a condition (or conditions) connected with a given path, an additional condition which has to be met to pass a given path (e.g. when special conversion between application is needed) and, finally, a list of input or output files for applications.

An expertise from a particular discipline is necessary to define rules for DMS. It can be done by a laboratory administrator together with a domain expert.

The DMS schema must be designed by the VLab administrator before it is available for users. Creating a new Dynamic Measurement Scenario needs a special attitude. We assume that some stages will be performed by the computer science engineer and scientist from a scientific discipline DMS concerns, and some by application.

The designing of the DMS consists of the following stages:

- application analyzing,
- preparing the connection diagram,
- describing additional dependencies in the connection diagram,
- generating applications and links description, and
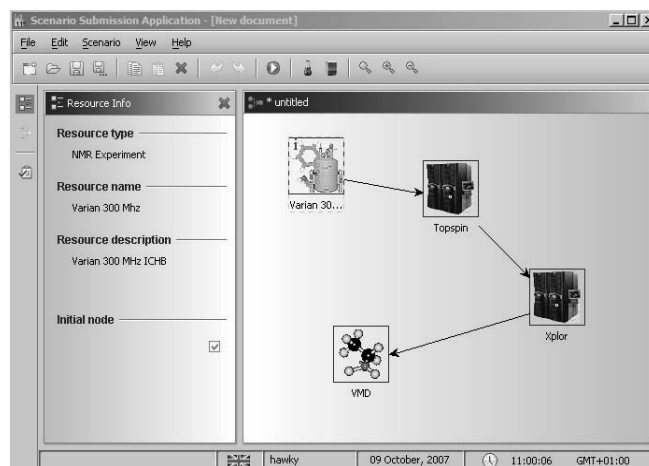- describing the measurement scenario.

Fig. 2. An example workflow in Scenario Submission Application

All stages are tightly connected between themselves and should be executed in the presented sequence. The first three phases should be performed, as we mentioned before, by the VLab administrator and engaged scientist. The last two can be done by a special application which take into consideration user's rights and information introduced by the user. Next, we will analyse each stage of DMS designing.

II.4. Digital Library

The crucial component in most typical RIS (and VL) systems is a module responsible for data storage and management (DSM). Performing remote experiments on laboratory devices in a multi-user environment is related to the necessity of storing results data which could be next used on the post-processing and visualization stage. This requirement is especially true for experiments where a certain number of devices need to work in parallel, e.g. in e-VLBI experiments. Execution of an e-VLBI experiment with 16 radio telescopes takes a few dozen of terabytes (dependently on measurement resolution). It illustrates the scale of the problem.

The functional requirements which DSM systems should meet are not limited to the storage and management. These systems should provide their users with a possibility to publish electronic papers and present the digital content.

Moreover, nowadays users want to have the possibility of using more and more sophisticated tools designed to create, browse and search for electronic documents. These tools perform an important role in the global information infrastructure, and can benefit to the education and scientific environments. These kinds of tools are called Digital Science Libraries (DSL).

DSL, which is implemented in the VLab system, is created on the base of the Data Management System (DMS) [1], which was developed for the PROGRESS project [2]. Its main functionality, which is storing and presenting data in grid environments, was extended with the functions specific to the requirements of the Virtual Laboratory [3].

This work presents the general assumptions, project design and implementation of the Digital Science Library for the exemplary purpose of Nuclear Magnetic Resonance Spectroscopy (NMR) data. The Digital Science Library should allow to store and present data used by the virtual laboratories. This data can be input information used for scientific experiments as well as the results of performed experiments. Another important aspect is the capability of storing various types of publications and documents which are often created in the scientific process. This type of functionality, which is well known to all digital library users, is also provided by the DSL.

The main functional assumptions of the Digital Science Library can be described in reference to the reasons for founding the Digital Library of Wielkopolska (Wielkopolska Biblioteka Cyfrowa – WBC) [10], developed by Poznań Supercomputing and Networking Center. Some of the most important assumptions are that:

- digital collections should be unique;
- digital library software should make it possible to significantly extend its usefulness;
- digital library software must cooperate with the library integrated systems and, in particular, allow using the catalogue database.

Referring to the first assumption, the Digital Science Library will serve the students and scientists by storing the results and computations data of the experiment. There is a need to digitalize the research in a form further called a digital publication.

The second assumption is related to the usefulness that should be provided by the digital library in comparison with the existing libraries. A digital library is not supposed to constitute a separate unit or to be in competition with the existing libraries. It should complete and broaden the overall library usefulness, expanding its universality and availability on the one hand, and on the other – allowing the librarian (or library administrator) to adapt the digital

library functionality to its readers (users) and administrator's needs.

In this context, DSL exemplifies the digital library which, in assumption, will serve the scientific unit providing the selected instrument (e.g. NMR spectrometer) through the Virtual Laboratory. Scientists should have a possibility of collecting and accessing results of their experiments and computations. The collected resources related to a defined institution and the provided instrument should have a teaching character, too.

The third issue concerns a semantic description of the stored resources, i.e. metadata. It allows searching the content of the library. DSL realizes these assumptions in a complete manner – the publications can be described by any metadata set, including these renowned as quasi-standards (e.g. Dublin Core). The issue of the scientists' experiment results is similar; in this case the available functionality is broadened by the possibility of storing the NMR data. For this purpose the NMR dedicated database described in point 5 has been developed.

To sum up, the Digital Science Library is an example of a digital library dedicated to scientists and offering the possibility to store the results of experiments and related documents. It should be outlined that, unlike another libraries, e.g. the Digital Library of Wielkopolska [10], DSL implements a different approach to the representation of digital resources. In WBC an electronic publication represents a separate content (an article, a report, a movie, a book), which can exist in the system in many versions, and the structure of that publication has a form of a single file (e.g. PDF) or multi-files (e.g. HTML). In DSL the electronic publication (an article, results of scientific work, a chemical compound, etc.) has a multi-file format and the metadata sets are used to model the possible relations and dependencies.

II.4.1. The Functionality of the Digital Libraries

Over the last several years, a few fundamental features which the modern digital library should provide [11] have been established. The list below presents the most important ones:

- Categorization – the ability to group the related publications;
- Catalogue description – detailed information about the digital entries in the library;
- Searching – the possibility of looking up the library content with the user-defined criteria (paper search or the experiment results search);
- Browsing – visualization of the search results in the user-friendly interface.

II.4.2. *Convenient and widespread access to the publications*

There are many scientists who find it necessary to publish their work results, materials, etc. in the digital library and share them with other users. However, to make it possible, the basic functionality described in the previous chapter is not sufficient. The Digital Science Library software has to be equipped with some additional features adequately to the requirements of the scientific environment. The list of such additions for the publication of documents and experiment results is presented below:

- Widespread access to the published papers – Internet access to the digital library resources has to be assured;
- Lastingness – a publication made available should never change its form (i.e. the file format) or localization (i.e. network localization);
- Version management – the possibility of keeping the document in many versions and managing them;
- Access management – resource access protection. The access cannot be granted to the valuable, non-published resources or experiment results;
- Copying protection – protection from unauthorized resource copying;
- Notifications – the user should be notified about the changes of the library content;
- Credibility – the published document should not be changed by non-trusted or unauthorized users.

Many existing digital libraries (e.g. WBC) can be characterized by the common qualities described above.

The Digital Science Library is built on the basis of the Data Management System. This implies the list of features which are not so common in the digital libraries environments but, on the other hand, are valuable and significant to the scientific environment. DMS is a distributed virtual file system, extended by the data storing on the Hierarchical Storage Management (HSM).

II.4.3. The Example Diagrams

After gathering all data which are designed for putting them into a database, the use case diagrams should be prepared for most frequent scenarios. One of them is discussed below.

There we can distinguish the main actor on the base of demands defined for the Digital Science Library of Nuclear Magnetic Resonance Spectroscopy. It is presented in Fig. 3. As an actor we can consider the system placed in the presentation or the indirect (middle) layer, on condition that it is able to communicate over the SOAP protocol.

The following diagram presents the two actors. The first – SOAPCompliedClient – is identified with a certain pattern of actor with granted rights for executing the specified operations, and provided with the ability of inter-communication with the library over the SOAP protocol. The nuSOAP (nuSOAP_PHPClient) is the second actor in the diagram. It can be used to create an access service for the NMR Library. In this case the access service is based on WWW pages and the PHP programming language.
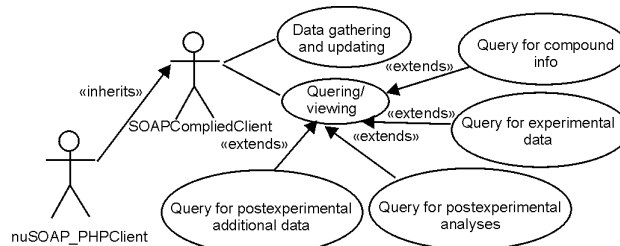

Fig. 3. Use case diagram

The diagram shows some main examples of using the DSL-NMR library system:

- Data gathering and updating – enables introducing and updating information related to the chemical compound and the experiment data.
- Querying / viewing – enables retrieving information on a chemical compound, experiment and its results, and the analysis executed on the basis of that data. The type of searching out the data depends on the needs. A standard list of requests consists of:

- query on compound info – condition: existence of information concerning re-searched chemical compound; searching out its attribute;
- query on experiment data – condition: existence of experiment data; searching out the attributes of an experiment, e.g. type of spectrum, sequence of impulses, spectrometer type, solvent, measurement temperature;
- query on post-experimental analysis – condition: existence of the analysis description; searching out of a chemical compound according to the value of chemical shift, coupling constant; and
- query on post-experimental additional data – condition: existence of the research documentation, e.g. reference on authors of experiments, for research papers documenting the received results.

III. GENOMICS

Genomics is one of the disciplines which can definitely benefit from the RIS system.Rapid development of genomics observed within the last 10 years was driven by

elaboration of new technologies, especially DNA sequencing and microarray technologies. Contrary to genetics, which focuses on individual genes, genomics concentrates on the whole genomes. Consequently, it attempts to answer the questions concerning the entire genome, e.g.: what is the size and content of the genome, how it is regulated, or how the expression of a particular gene influences all others genes, etc. Although all cells forming an organism contain the same genetic information, their size, shape and function can be distinctly different. It is possible since in each kind of cells a different set of genes is expressed. To release the information encoded in dsDNA, individual genes are transcribed into RNA which then serves as a template for the protein synthesis. Accordingly, the present status of any cell can be determined by analyzing the overall RNA or protein content, called transcriptome and proteome, respectively. Thus, by applying DNA microarrays or other techniques used in functional genomic, one can precisely distinguish cells coming from different tissues, cells being at the different stage of development, or cells derived from healthy and pathologically changed organs.

III.1. Microarray Experiment Execution

The DNA microarray is a set of DNA molecules called probes, spotted on a solid surface, for example glass slide, in an ordered fashion [16, 17]. Each probe is complementary to a specific fragment of one gene. At present it is possible to manufacture DNA microarrays containing over six million probes located at 1 cm square [14]. Consequently, the expression of all genes can by analyzed simultaneously with a single microarray. At present many companies offer standardized, ready to use microarrays [14, 15]. However, they are rather expensive and not available for each organism, especially when its genome is not sequenced. Alternatively, it is possible to design probes and produce home-made microarray, e.g. dedicated to a particular research project. This approach is more flexible and economic.

A microarray experiment scheme is presented in Fig. 4. Construction of the microarray is a separate step that can be avoided when we use commercial microarrays. Another challenging task is preparing a sample to study with a microarray. First, all mRNA molecules (primary products of gene expression) are extracted from the target samples. Then mRNAs are converted into complementary single-stranded DNA molecules (cDNA) in the process called reverse transcription. During this stage cDNAs are also labeled with fluorescent dyes (e.g. cyanines). Sometimes, when the amount of biological material is very limited, cDNA amplification is required.
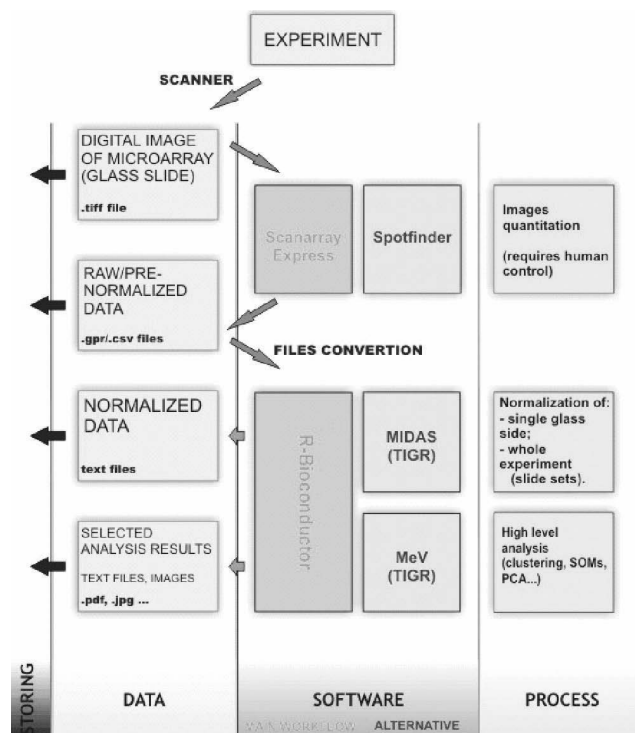

Fig. 4. Microarray experiment scheme

In case of a two-color experiment, two samples – the control and the studied one – are prepared simultaneously. Each sample is labeled with a different dye (e.g. the control sample with Cy3, the examined sample with Cy5). Both samples are subsequently mixed and incubated overnight with the same microarray. During this stage, fluorescently tagged cDNAs hybridize with complementary probes. Theoretically, each type of cDNA should only hybridize with one specific probe. Then the microarray is washed, dried and scanned with two lasers. Each of them selectively detects signals coming from one of the applied fluorescent dyes. The intensity of the detected signal is directly proportional to the number of hybridized cDNA molecules, thus it also reflects the level of corresponding gene expression. As a result, two images are generated, the first showing the level of gene expression in the tested and the second in the control sample. The differences in the activity of the individual genes as well as the whole genomes can be superficially determined by superimposing of both images. Usually, however, a more detailed statistical analysis is required. During the quantitative analysis images are converted to the numerical text data file. Raw data are subsequently submitted to the pre-processing – normalization procedures – and so-called low-level analysis, e.g. gene filtration, detection of differential expression. Normalization procedures minimize discrepancies between

slides and allow for comparison of the results obtained from different experiments – high-level analysis, encompassing clustering of the samples and probes, correlation network modeling, construction of gene ontology trees and a biochemical pathway, etc. [20-24]. Apart from many commercial programs dedicated to the analysis of micorarray data and less abundant free software, the most popular and flexible platform is Bioconductor in the R environment [12]. Bioconductor is a still developing software project, offering a rich collection of tools (packages) useful for the analysis and comprehension of genomic data. The results of an analysis performed in Bioconductor can be deposited as pdf, jpg or bmp reports.

III.2. Genomic Virtual Laboratory

There are numerous examples that DNA microarrays can be very useful in both basic and applied research, especially in medicine. It was demonstrated that the DNA microarray-based analysis of medical samples allows doctors for fast and accurate diagnostics, selection of the optimal treatment method and reliable prediction of the results of applied therapy [25, 18, 19]. Unfortunately, common usage of DNA microarrays-based techniques is currently rather difficult. There are at least two major limitations: the cost of the sophisticated equipment and a small number of appropriately educated specialists. Thus, two possible solutions can be applied in order to overcome the difficulties:

- specialized genomic centers, and
- virtual laboratories.

Genomics centers are still more popular. In the USA and Western Europe such centers often function as integral parts of clinics, conducting the research projects as well as providing diagnostic services. In Poznań there is a Regional Genomics Center (RGC), a consortium founded by the Institute of Bioorganic Chemistry, University of Medical Sciences and University of Technology in Poznań. It is a unique Polish laboratory equipped with a complete set of instruments necessary for the production, hybridization, scanning and analysis of the commercial or home-made microarrays. Thanks to collaboration of scientists representing different branches of science (biology, biotechnology, medicine and informatics), more complex and professional analyses can be performed there.

The genomic virtual laboratory is a better solution than a specialized genomics center as it provides common and parallel access for users that cannot directly collaborate with a specialized center. Theoretically, it should be possible to prepare the sample, send it by post and have online control the instruments and programs that can be used for

analysis of this sample. Unfortunately, not every step of a microarray experiment can be fully automated – some instruments still need to be operated manually. Probably more microarray tools will be web-available in the future.

Nowadays, at least computational work connected with microarray probes design and analysis of microarray results can be automated. The rest of the experimental procedure has to be performed by a specialist. For example, a hospital which is not equipped with sophisticated devices and does not employ specialists trained in microarray analysis, can send a medical sample to a virtual laboratory where the sample will be processed, and then monitor the process of a microarray-based analysis. In the end the obtained results can be confronted with the analogical data generated earlier with the samples of the same type.
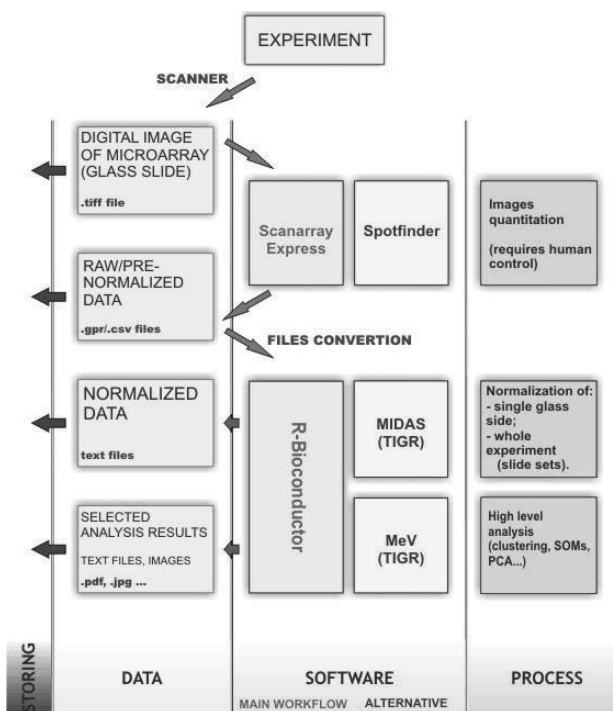

Fig. 5. Diagram presenting experiment execution in virtual laboratory of genomics

As presented in Fig. 5, after scanning a microarray image (deposited in a tiff file) is quantitated using e.g. Spotfinder or commercially available programs – ScanArray Express, Spot, GenePix, ImaGene etc. Raw data in a text file (csv, gpr or similar) are then submitted to normalization, filtration and high-level analysis, using sophisticated software (e.g. R Bioconductor, Multi Experiment Viewer, Gene Spring, Genowiz). Result reports are saved as text or graphic files (e.g. pdf, jpg). Data transfer within programs

usually demands file conversion. For example, gpr files generated by ScanArray Express do not follow the GenePix standards and are not automatically recognized by Bioconductor or by MIDAS (TIGR). Similarly, an output file generated by MIDAS is recognized by related MeV (Multi Experiment Viewer, TIGR) but not by Bioconductor.

III.3. Experiment Description

Measurement in a virtual laboratory starts from the post-processing stage. A microarray image can be stored in a digital library and next processed by particular applications according to the workflow diagram (Fig. 6).
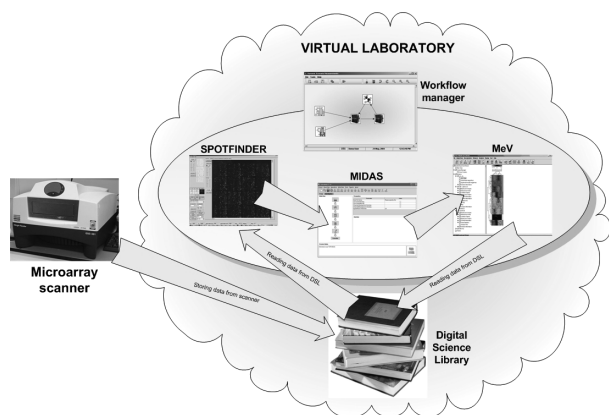

Fig. 6. Outlook of Virtual Laboratory of Genomics

The added value in this solution is automation of the quantitation procedure. The end user can focus only on his/her research field – analysis – and draw conclusions from genomic data. All technical aspects related to server and software installation, creation of a new user account, setting up access rights, etc., are hidden from users. Another positive aspect is the possibility of storing all result data in a digital science library in an organized way. Output data from each stage are stored in it. It gives a possibility to go back in the experiment procedure and repeat any phase of research. Moreover, users can share experiment data among them and work together on a given project.

Finally, what is worth highlighting is an aspect of simplicity in a virtual laboratory. The end user operates only on graphical applications with an intuitive interface. It could be very important especially for people without any technical background.

IV. SUMMARY

Remote Instrumentation Systems are becoming more and more popular among users wishing to use very rare and expensive scientific devices remotely. These systems, apart from remote access, provide an environment which facilitates experiments conducted on many stages.

Here, we have presented the general presumption of the virtual laboratory, showed how it was applied in an NMR field, and discussed the implementation of this idea in functional genomics. The main advantage of this system will be supporting users in a widely understood post-processing phase. It is assumed that processing scenarios are fully automated and aided by virtual laboratory software. One of the most important modules in this system is a workflow management tool which allows to define the analysis path. Another one is the Digital Science Library used for output data storing and management. The DSL system also supports teamwork by sharing experiments results among different groups of users.

The purpose of future work will be developing a detailed project of the Virtual Laboratory of Genomics system which will support and automate all activities related to processing data after the measurement phase.

Acknowledgments

References

[1] *PROGRESS – Polish Research on Grid Environment for SUN Servers.* http://progress.psnc.pl/English/index.html.
[2] *Virtual Laboratory PSNC.* http://vlab.psnc.pl/ .
[3] B. R. Seavey, E. A. Farr, W. M. Westler and J. L. Markley, *A Relational Database for Sequence-Specific Protein NMR Data.* J. Biomolecular NMR **1**, 217-236 (1991).
[4] *NMR data-sets Bank: A repository for raw NMR data-sets.* http://nmrb.cbs.cnrs.fr/index.html.
[5] Ch. Steinbeck, S. Krause and S. Kuhn, *NMRShiftDB – Constructing a Free Chemical Information System with Open-Source Components.* J. Chem. Inf. Comput. Sci. **43**, 1733-1739 (2003).
[6] O. Yamamoto, K. Someno, N. Wasada, J. Hiraishi, K. Hayamizu, K. Tanabe, T. Tamura and M. Yanagisawa, *An Integrated Spectral Data Base System Including IR, MS, 1H-NMR, 13C-NMR, ESR and Raman Spectra.* Anal. Sci. **4**, 233-239 (1988).
[7] The National Science Digital Library, http://nsdl.org/ .
[8] Digital Library for Earth System Education, http://www.dlese.org/ .
[9] Poznań Supercomputing and Networking Center, http://www.man.poznan.pl.
[10] *Globus Toolkit.* http://www.globus.org/toolkit/ .
[11] *GRMS,* http://www.gridlab.org/WorkPackages/wp-9/ .
[12] *Bioconductor project.* http://www.bioconductor.org/.
[13] *TIGR TM4 Suite* (including Spotfinder, MIDAS, MeV). http://www.tm4.org/.

[14] D. Gershon, *More than gene expression.* Nature **437**, 1195--1198 (2005).

[15] D. N. Howbrook, A. M. van der Valk, M. C. O'Shaugnessy, D. K. Sarker, S. C. Baker and A. W. Lloyd, *Developments in microarray technologies.* Drug Discov. Today **8** (**14**), 642-651 (2003).

[16] S. Venkatasubbarao, *Microarrays – status and prospects.* Trends Biotechnol. **22**, 630-637 (2004).

[17] V. Trevino, F. Falciani and H. A. Barrera-Saldaña, *DNA microarrays: a powerful genomic tool for biomedical and clinical research.* Mol. Med. **13**, 527-541 (2007).

[18] O. Margalit, R. Somech, N. Amariglio and G. Rechavi, *Microarray-based gene expression profiling of hematologic malignancies: basic concepts and clinical applications.* Blood Rev. **19**, 223-234 (2005).

[19] T. Haferlach, A. Kohlmann, S. Schnittger, M. Dugas, W. Hiddemann, W. Kern and C. Schoch, *Global approach to the diagnosis of leukemia using gene expression profiling.* Blood **4**, 1189-1198 (2005).

[20] Y. H. Yang, S. Dudoit, P. Luu, D. M. Lin, V. Peng, J. Ngai and T. P. Speed, *Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variations.* Nucl. Acids Res. **4**, e15 (2002).

[21] G. K. Smyth and T. P. Speed, *Normalization of cDNA microarray data.* Methods **31**, 265-273 (2003).

[22] J. Quackenbush, *Microarray data normalization and transformation.* Nature Genet. Suppl. **32**, 496-501 (2002).

[23] G. K. Smyth, Y. H Yang and T. Speed, *Statistical issues in cDNA microarray data analysis.* Methods Mol Biol. **224**, 111-36 (2003).

[24] B. I. P. Rubinstein, McAuliffe, S. Cawley, M. Palaniswami, K. Ramamohanarao and T. P. Speed, *Machine Learning in Low-level Microarray Analysis.* ACM SIGKDD Explor. Newsletters **5 (2)**, m14 (2003).

[25] L. A. Garraway and W. R. Sellers, *Array-based approaches to cancer genome analysis.* Drug Discov. Today **2 (2)**, 171--177 (2005).

LUIZA HANDSCHUH graduated in biology from the Adam Mickiewicz University in Poznań in 1998. She received her PhD degree from the Institute of Bioorganic Chemistry in 2004. Since 2005 she has been a specialist in the Department of Haematology, Poznań University of Medical Sciences, and since 2006 – a research associate in the Center of Excellence for Nucleic Acid-based Technologies, one of the laboratories affiliated to the Institute of Bioorganic Chemistry in Poznań. Her main interest is functional genomics, DNA microarray technology and its application in medical science. She participates in many research projects, collaborating with other Polish laboratories. At present, she is one of the principal investigators in the project devoted to gene expression profiling in a human acute myeloid leukemia.

MARCIN LAWENDA graduated from the Poznań University of Technology and received M.Sc. in Computer Science (Parallel and Distributed Computation area) in 2000. He defended the Ph.D. thesis on the same university at Faculty of Computing Science and Management, Institute of Computing Science in 2006. He is currently working for Poznań Supercomputing and Networking Center on the project manager position on the Virtual Laboratory project. His research interests include parallel and distributed computing, scheduling and Grid technologies. He has been a member of the Polish Information Processing Society since 2000.

PIOTR STĘPNIAK received M.Sc. in Biotechnology in 2008 from Adam Mickiewicz University in Poznań. He has been studying microarray technology in the Institute of Bioorganic Chemistry, Polish Academy of Sciences since 2006. His thesis focused on microarray data analysis. He has been helping to develop the Virtual Laboratory of Genomics providing R-environment-based data analysis solutions. He is currently a Ph.D. student in the BioInfoBank Institute in Poznań, expanding his knowledge and skills in the field of bioinformatics.

MAREK FIGLEROWICZ (MF), graduated in chemistry from the Poznań Institute of Technology in 1985. He received the Ph.D. degree in biochemistry from the Adam Mickiewicz University in 1991 and the D.Sc. (habilitation) degree in biology from the same university in 2000. In 2006 he acquired the title of professor of biology from the President of Poland. His research interests concern: RNA biology, especially small regulatory RNA; molecular biology of RNA viruses and retroviruses, functional genomics. He has been the principal investigator of numerous scientific project sponsored by the Polish Ministry of Science and High Education, European Union and UNESCO. He is the author or co-author of 91 articles and 122 published scientific communications. As an invited speaker, he has given 47 lectures and seminars. Currently, he is the head of the Laboratory of Plant Molecular Biology at the Institute of Bioorganic Chemistry (IBCH) Polish Academy of Sciences, Poznań, Poland, the head of the Center of Excellence for Nucleic Acid-based Technologies (CENAT), located at the IBCH and the head of the Laboratory of Molecular Virology at IBCH. He is also a member of several scientific boards, the editorial board of Journal of RNAi and Gene Silencing, Oxford United Kingdom and a member of the editorial board of Biotechnologia – Journal of the Biotechnology Committee of the Polish Academy of Sciences.

DR. MACIEJ STROIŃSKI is the Technical Director of the Poznań Supercomputing and Networking Center, Vice President of the PIONIER Consortium – Polish Optical Internet – and member of the Supervisory Board. He is also lecturer in the Institute of Computing Science of the Poznań University of Technology. His research interests concern networking, quality of service, network protocols and management, and grid-related topics. He is author and co-author of over 150 papers in major professional journals and conference proceedings. Maciej Stroiński is the co-author of the Pionier Programme, accepted by the Polish State Committee for Scientific Research and announced as the official Polish programme in 2000. He is a member of Programme and Steering Committees of national and international conferences and editor of Computational Methods in Science and Technology journal. Maciej Stroiński participated in several networking and grid-related EU and national projects, co-operating with well known IT companies. He established Sun Center of Excellence and Microsoft Innovation Center in PSNC.

JAN WĘGLARZ, Academician, Professor (Ph.D. 1974, Dr. Habil. 1977), in years 1978-83 Associate Professor and then Professor in the Institute of Computing Science, Poznań University of Technology, member of the Polish Academy of Sciences (PAS), Director of the Institute of Computing Science, Poznań University of Technology and its predecessors since 1987, Director of Poznań Supercomputing and Networking Center, President and Scientific Secretary of the Poznań Branch of the PAS, Vice President of the Committee for Computer Science of the PAS, member of the State Committee for Scientific Research, Principal Editor of the Foundations of Computing and Decision Sciences, member of several editorial boards, among others Internat. Trans. Opnl. Res. and European J. Opnl. Res. Representative of Poland in the Board of Representatives of IFORS and in EURO Council (President of EURO in years 1997-98). Member of several professional and scientific societies, among others the American Mathematical Society and the Operations Research Society of America. Author and co-author of 11 monographs, 3 textbooks (3 editions each) and over 200 papers in major professional journals and conference proceedings. Frequent visitor in major research centers in Europe and the USA. Co-laureate of the State Award (1988) and the EURO Gold Medal (1991), laureate of the Foundation for Polish Science Award (2000).