

Capacities/Research Potential
FP7-REGPOT-2012-2013-1

Project No. 316254

BASTION

“From Basic to Translational Research in Oncology”

Deliverable D4.2

**Report on creation of IT infrastructure for supporting
personalized medicine approaches in oncology**

Project start date:	1.09.2012
Project duration:	42 M
Due date of deliverable:	28.02.2014
Actual submission date:	18.03.2014
Dissemination level:	PUBLIC

Table of content

Introduction	3
1. Brief description of the designed IT infrastructure	4
2. Components of the IT infrastructure	5
2.1. The computing cluster.....	5
2.2. The workstations	10
2.3. The computer network	12
2.4. The software.....	14
3. An Uninterruptible Power Supply	15
Conclusions	17
Corresponding estimated budget	18
Attachment: Functional diagram of the BioInfo Laboratory Infrastructure.....	19

All reports are available on BASTION Webpage: www.bastion.wum.edu.pl

Introduction

Deliverable D4.2 corresponds to the T4.7 task that was delivered with 15 days delay. The delay were due to an unexpected delay in the delivery of 4 SSD drivers, which were components of four work stations (the latter delivered on time).

This deliverable regards report on designing and constructing an IT infrastructure for bioinformatics, supporting in particular personalized medicine in oncology. The modern medical sciences relay heavily on computing and data storage platforms. High throughput technologies provide researchers with the means of experimental analysis on the scale unthinkable just ten years ago. On the other hand, the volume of data generated by such experiments poses heavy requirements on computing infrastructure.

The IT infrastructure built within the BASTION project offers the Medical University of Warsaw's researchers an universal platform for undertaking ambitious studies in the area of bioinformatics. The solution was carefully designed to meet several, often conflicting, goals: high performance, extensive storage capacity, top data security and maximum usage flexibility.

Because the allocated budget was tightly constrained, the IT infrastructure design process required extensive market research and careful fund allocation among the elements of the IT system. After the detailed audit of the existing infrastructure it became clear that, in order to come up with the satisfying solution, the originally allocated funds might not be sufficient, posing a serious risk of scarifying some vital parameters of the IT system. Therefore, taking into account the importance of this task for the whole BASTION project, it was decided that the purchase will be delayed, hoping to utilize the funds better, as the price-to-performance ratio of computer equipment usually lowers with time. On the top of that, the deployment of a computing cluster required installation of a new air-conditioning equipment for the computing cluster server room and an adaptation of the electric circuits powering it. All of the above-mentioned circumstances had a profound effect on the final timing of the purchase procedures.

Last but not least, one of the criteria for the design was the infrastructure energy use efficiency. As a result, most of the elements of the created infrastructure are characterised by the high performance-to-energy-consumption ratio. It will further translate into lesser environmental burden and operational cost savings.

Table I presents the summary of the purchased elements of the IT infrastructure

Table I. Summary of the purchased elements of the IT infrastructure

No	Task	Equipment	Agreement no	Date of signing the agreement	Date of installation /purchase	Invoice delivery date	Net price [EUR]
EQUIPMENT							
1	T4.7.1	Computing cluster - mass storage subsystem, including back-up solution	AEZ/365/S-299/32/40/2014	2014.01.31	2014.02.28	2014.02.28	86 616
2	T4.7.2	Computing cluster - computing servers	AEZ/365/S-299/13/41/2014	2014.01.31	2014.02.28	2014.02.28	44 080
3	T4.7.3	Multicore workstations (with WQHD monitor screens)	AEZ/365/S-299/13/43/2014	2014.02.06	2014.03.06 (SSD drivers 2014.03.14)	2014.03.18	11 503
SOFTWARE							
4	T4.7.4	Specialised software for data analysis and visualisation	An order: ATZ_AP_1M19_2013/EL/13104/2013	-	2014.03.14/2013.12.20	2014.01.07	2 393
OTHER COSTS FOR T4.7							
5	T4.7	Computing cluster server room air-conditioning equipment	AEZ/365/S-014/076/2014	2014.02.24	2014.02.28	2014.02.28	3 230
6	T4.7	Computer network equipment	AEZ/365/S-299/13/44/2014	2014.01.31	2014.02.19	2014.02.25	11 215
OTHER COSTS FOR T4.2-T4.3							
7	T4.2-T4.3	UPS	AEZ/365/S-299/13/42/2014	2014.01.31	2014.02.21	2014.02.27	5 615

1. Brief description of the designed IT infrastructure

The system as a whole has been designed to provide the researchers of the Medical University of Warsaw with the means to successfully tackle various challenges regarding gathering and analysis of the biological and medical data. The ultimate goal was to come up with an universal, balanced and flexible solution possessing multiple strengths.

The resulting system makes possible to efficiently store, process and analyse complex data from the genomic, transcriptomic and proteomic experiments. It can also handle 3D structure modelling supporting drug discovery research and, in a foreseeable future, can also act as a support for digital pathology solutions.

The network infrastructure for the new computing cluster has been designed in a way allowing for its efficient integration with already existing massive computing power offered by the nearby facilities of the Interdisciplinary Centre for Mathematical and Computational Modelling (<http://www.icm.edu.pl>). Attachment to the deliverable shows the functional diagram of the BioInfo Laboratory infrastructure.

2. Components of the IT infrastructure

An exemplary solution of an IT infrastructure that the created system was based upon, consists of the following elements:

1. a computing cluster,
2. four workstations,
3. a computer network.

The hardware part of the system forms a foundation for executing software to deal with actual bioinformatics challenges. Each of the aforementioned elements is discussed below.

2.1. The computing cluster

The cluster includes:

1. a mass storage subsystem with a back-up solution,
2. computing servers,
3. server room infrastructure.

Purchased equipment:

- **Computing cluster mass storage subsystem, including back-up solution**, purchased by MUW from “Slash Sp. z o. o.” company; major elements of the purchase:

- one Dell PowerVault NX3610 NAS appliance (with two controllers) complemented with two Dell PowerVault MD3660i iSCSI arrays (with two controllers each),
- two Dell PowerConnect 8132 LAN switches,
- two Dell PowerEdge R720 servers,
- Rack infrastructure, including one Dell 42U rack cabinet and two Dell UPS 2700W uninterruptible power supplies,

- **Computing servers** purchased by MUW from “Slash Sp. z o. o.” company; major elements of the purchase:

- six Dell PowerEdge R620 servers, each with four Intel DC S3700 200GB solid state drives,
- one Dell PowerConnect 2816 LAN switch;

- **Computing cluster server room air-conditioning equipment**, purchased by MUW from “KAWER Dominik Werker” company; major elements of the purchase:
 - two Fujitsu inverter air-conditioning sets (ASYG18LFCA+AOYG18LFC+ a wired remote controller per each set).

Volume of raw data coming out of high-throughput experiments is often high. For example, a single Next Generation Sequencing experiment carried out at the Medical University of Warsaw generates hundreds of gigabytes of data. Because of this massive volume, especially when one considers analysing many data sets at the same time, the storage platform must feature high transfer rates and equally fast network connections. As the experiments are costly and often impossible to repeat because of the limited availability of the unique biological material, stored data have to be properly backed-up. In order to avoid unexpected down-times the solution should be modular and redundant, based on hot-swappable elements.

The data analysis aspects are at least as demanding as data storage challenges. To efficiently cope with the sheer volume of the high throughput experimental data one needs a high computing power. It is offered by the purchased computing cluster, which is in the centre of the system.

Brief technical specification of:

Computing cluster - mass storage subsystem, including back-up solution

Parameter	Requirement(s)
raw storage capacity	320TB
storage protocol support	<ul style="list-style-type: none"> • iSCSI • NFS • CIFS
data security	<ul style="list-style-type: none"> • redundant and hot-swappable: disks, controllers, power supplies • safe write cache solution • built-in integrity tests • remote error reporting
LAN connectivity	10GBASE-T

Date of the tender announcement: 3rd January, 2014
 Date of the tender announcement decision (outcome): 24th January, 2014
 Date of signing the AEZ/365/S-299/13/40/2014 agreement: 31st January , 2014
 Remaining offers (valid ones ranked according to their price):

- “Zakład Systemów Komputerowych ZSK Sp. z o . o.” company – invalid offer, missing required documentation

Date of installation: 21st-28th February, 2014

Date of user training:

28th February, 2014

Trained persons:

No.	Name	Department
1	Sławomir Gruca	Department of Immunology, WUM
2	Piotr Stawiński	Department of Immunology, WUM

Brief technical specification of:

Computing cluster - computing servers (all combined)

Parameter	Requirement(s)
computing power	<ul style="list-style-type: none"> at least 150000 PassMark® CPU Mark points
RAM size per thread	<ul style="list-style-type: none"> at least 8GB
RAM throughput per thread	<ul style="list-style-type: none"> at least 3.4GB/s
SSD raw storage capacity	<ul style="list-style-type: none"> at least 4800 GB at least 10WPD 5-year-long endurance write cache power loss protected
LAN connectivity	10GBASE-T

Date of the tender announcement:

3rd January, 2014

Date of the tender announcement decision (outcome):

24th January, 2014

Date of signing the AEZ/365/S-299/13/41/2014 agreement:

31st January, 2014

Remaining offers (valid ones ranked according to their price):

- “Alternative Technology Sp. z o. o.” company – invalid offer, missing required documentation
- “Zakład Systemów Komputerowych ZSK Sp. z o. o.” company – invalid offer, missing required documentation
- “Spinel Sp. z o. o.” company - invalid offer, missing critical information

Date of installation:

25th-28th February, 2014

Date of user training:

28th February, 2014

Trained persons:

No.	Name	Department
1	Sławomir Gruca	Department of Immunology, WUM
2	Piotr Stawiński	Department of Immunology, WUM

Brief technical specification of:

Computing cluster server room air-conditioning equipment:

Parameter	Requirement(s)
nominal cooling power (for each of the two sets)	in the range from 5 to 7 kW
type	inverter, split-system
ERR	at least 3.4
controller	wired

Date of the tender announcement: 15th January, 2014

Date of the tender announcement decision (outcome): 14th January, 2014

Date of signing the AEZ/365/S-014/076/2014 agreement: 24th February, 2014

Remaining offers (valid ones ranked according to price 85% and warranty period 15%):

- “ERCO Przemysław Ryszard Żelewski” company – invalid offer, offered installation time exceeded the required
- “ENERGY HEAT TECHNOLOGY INVESTMENT Burcan Marek” company – 64,64% total
- “PPHU PRZEMKO Przemysław Biedrzycki” company – 88,41% total
- “FAVORIT Sp. z o. o.” company – invalid offer, missing required documentation
- “ETIS Sp. z o. o.” company – invalid offer, offered installation time exceeded the required
- “Thermodynamika s.c. M.Janus, P.Jarocki” company – invalid offer, required technical parameter(s) not met
- “Klim-Spaw Sp. z o. o.” company – invalid offer, offered installation time exceeded the required

Date of installation: 28th February, 2014

Date of user training: 28th February, 2014

Trained persons:

No.	Name	Department
1	Slawomir Gruca - IT Professional	Department of Immunology, WUM
2	Piotr Stawinski - IT Professional	Department of Immunology, WUM

Equipment use:

This equipment will be widely used by the researches of the Medical University of Warsaw. It will provide convenient, efficient and safe storage space not only for the storage of the experimental data but also for vital documents and reports.

As the BASTION researchers plan to establish a digital pathology facility in the nearest future, massive quantities of imaging data are likely to be generated and make heavy use of

the storage space. The available mass storage space shall enable to store thousands of virtual scans and provide quick access to them.

The computing cluster will ease the bottleneck of processing the data of the Next Generation Sequencing experiments, making the results available to the researchers more quickly. Moreover, more computationally expensive processing will be possible, allowing for more advanced and accurate data analysis.

In order to provide the computing equipment with proper environmental conditions, a new air-conditioning equipment had to be purchased and installed in a server room. It is a fully redundant solution, which minimizes IT hardware failure odds; the air-conditioning system is critical for proper functioning of the system.



Photo 1 (left): The Rack cabinet containing elements of the computing cluster, namely mass storage devices and computing servers. This hardware is located in the server room of Department of Medical Genetics, MUW.



Photo 2 (right): The internals of the cabinet of the computing cluster: six computing servers at the top and mass storage devices in the lower part.



Photo 3: The heat generated by the Computing cluster in the server room is evacuated by two air-conditioning equipment sets. The redundancy is vital, as it decreases the odds of a failure of the cooling system, which would result in rapid increase of the server room temperature and cause damage to the cluster equipment.

2.2. The workstations

Purchased equipment:

- **Four multicore workstations** (EQUIPMENT, T4.7.1), purchased by WUM from “Spinel Sp. z o. o.” company:
 - four Hewlett-Packard Z230 workstations, each equipped with two Hewlett-Packard OEM Samsung SM843T 240GB Solid State Drives and two Hewlett-Packard Z27i monitors.

The workstations for bioinformatics applications need to possess several qualities related to the specificity of the biodata. Its volume requires high-performing storage and network components. On the other hand, considering budget constraints and having at a disposal the computing cluster, investing in pure central processing unit (CPU) power of workstations would not be reasonable. Thus, the CPU requirements were eased, so that more balanced solution was achieved.

Brief technical specification (for each of the workstations):

Parameter	Requirement(s)
computing power	<ul style="list-style-type: none"> at least 9400 PassMark® CPU Mark points
RAM size	<ul style="list-style-type: none"> at least 32 GB
SSD storage size	<ul style="list-style-type: none"> at least 480 GB at least 10WPD 5-year-long endurance write cache power loss protected
LAN connectivity	10GBASE-T
monitors	27 inch 2560x1440 pixels

Date of the tender announcement:	3 rd January, 2014
Date of the tender announcement decision (outcome):	24 th January, 2014
Date of signing the AEZ/365/S-299/13/43/2014 agreement:	6 th February, 2014
Remaining offers (valid ones ranked according to their price):	
<ul style="list-style-type: none"> “Zakład Systemów Komputerowych ZSK Sp. z o. o.” company - invalid offer, missing required documentation “Slash Sp. z o. o.” company – 99,59% total 	
Date of the workstations purchase:	6 th March, 2014
Data of the missing SSD drives purchase:	14 th March, 2014

Equipment use:

The workstations will be utilized by researchers for highly resourced intensive bioinformatics tasks, like analysing of the “-omics” data. The purchased equipment is well tailored for such use as it is equipped with fast and reliable SSD storage and able to efficiently handle moving of the large files across computer network. The computing power parameters are also well above the average of today's personal computers.

What is worth mentioning, ergonomic and productivity aspects were also considered when planning the purchase, as the workstations are equipped with quality displays. Such monitors, apart from making the computer work more comfortable, provide an additional value - sometimes an eye inspection of properly visualized data yields valuable insights and results in the new discoveries. Moreover, the size and resolution of the screens will also be helpful when working with digital pathology applications.



Photo 4: One of the four bioinformatics workstations. It is equipped with two high resolution screens, SSD storage and a 10Gbps network interface

2.3. The computer network

Purchased equipment:

- **Computer network equipment** (OTHERS FOR T4.7), purchased by WUM from “System Data Sp. z o. o.” company:
 - four Hewlett-Packard 2920-48G LAN switches, equipped to support 10GBASE-T, 10GBASE-LR and 1000BASE-LX connectivity

A computer network is the foundation of virtually any computer system. After initial evaluation of an existing IT infrastructure, it was concluded that the network had to be upgraded. Speed of the links was severely limited, rendering the network practically unsuitable for sending large data sets. Therefore, a decision was made to purchase network equipment that would enable to take advantage of contemporary high-speed networking standards, thus removing the serious bottleneck from the system.

As most of the elements of the final solution, that is the computing cluster, mass storage, and several workstations, are equipped with high-speed network interfaces, having a proper network equipment will enable to move data even 100 times faster, transforming hours of waiting for data transfer to complete - into minutes. That has an enormous impact on user experience and productivity. Among other things, it obsoletes the need to use external hard drives as the only practical mean to transport large data volumes between laboratories.

Brief technical specification (for each of the switches):

Parameter	Requirement(s)
internal throughput	<ul style="list-style-type: none"> at least 175 Gbps
LAN connectivity	<ul style="list-style-type: none"> 10GBASE-T 1000BASE-T 10GBASE-LR 1000BASE-LX
IPv6	<ul style="list-style-type: none"> implemented, including management interface at least static routing supported

Date of the tender announcement: 3rd January, 2014
 Date of the tender announcement decision (outcome): 24th January, 2014
 Date of signing the AEZ/365/S-299/13/44/2014 agreement: 31st January , 2014
 Remaining offers (valid ones ranked according to their price):

- “KOMA NORD Sp. z o. o.“ company – 75,38%
- “Alternative Technology Sp. z o. o.“ company - invalid offer, missing required documentation
- “Spinel Sp. z o. o.“ company – 46,19%
- “B3System S.A.“ company - invalid offer, missing required documentation

Date of purchase: 19th February, 2014

Equipment use:

The network equipment is an integral part of the IT infrastructure as it integrates all the parts of the solution. The purchased network equipment will practically be used by many researches of the Medical University of Warsaw. It is actually difficult to underestimate the importance of a computer network for any IT infrastructure, not to mention the one for bioinformatics. It is the core of it and enables an efficient use of the computing and storage resources.



Photo 5: One of the four purchased LAN switches; the plastic transparent box on the top contains optical fiber communication modules.

2.4. The software

Purchased items:

- Specialised software for data analysis and visualisation (SOFTWARE, T4.7.4), purchased by WUM from “Oprogramowanie Naukowo-Techniczne” company:
 - two academic network licences of MathWorks® MATLAB®, each extended with:
 - Statistics Toolbox™ licence,
 - Bioinformatics Toolbox™ licence,
 - Image Processing Toolbox™ licence,
 - Database Toolbox™ licence.

To make sense of the experimental and medical data the computer hardware itself is not enough - it has to be complemented by the software for data processing, analysis and visualization. Fortunately, there are numerous software packages available today that fulfil those purposes. Many of them are free open-source software widely accepted in scientific community and the researchers of the Medical University of Warsaw heavily use those on a daily basis. Nevertheless, a proprietary specialised software products often are more convenient to use, rich in features and are well supported.

MathWorks® MATLAB® software is one of leading scientific specialised software suited for data analysis and visualization. Its flexible licensing allows for an efficient use of this software in the laboratory.

No tender procedure was required, only an order for a delivery.

Date of purchase:

20th December, 2013

Software use:

Although the cost of this purchase was a small fraction of the task's budget, the software will be used by the BASTION researchers for carrying variety of data processing, including “-omics” data analysis and visualization. Thanks to the software flexibility, solving a wide spectrum of future bioinformatics challenges shall be possible. Moreover, the software might become an important tool for the planned digital pathology facility, as it is capable of high-resolution image processing and analysis.

Comment on database management software

A proprietary database management software is usually expensive and involves significant annual payments. Its purchase could undermine the continuity of the created solution, after the project funding would have ended. As there are multiple open-source free-of-charge mature and advanced alternatives available, it would not be reasonable to buy proprietary and expensive software, especially when the available funds were scarce. Therefore it was decided to waive that particular purchase.

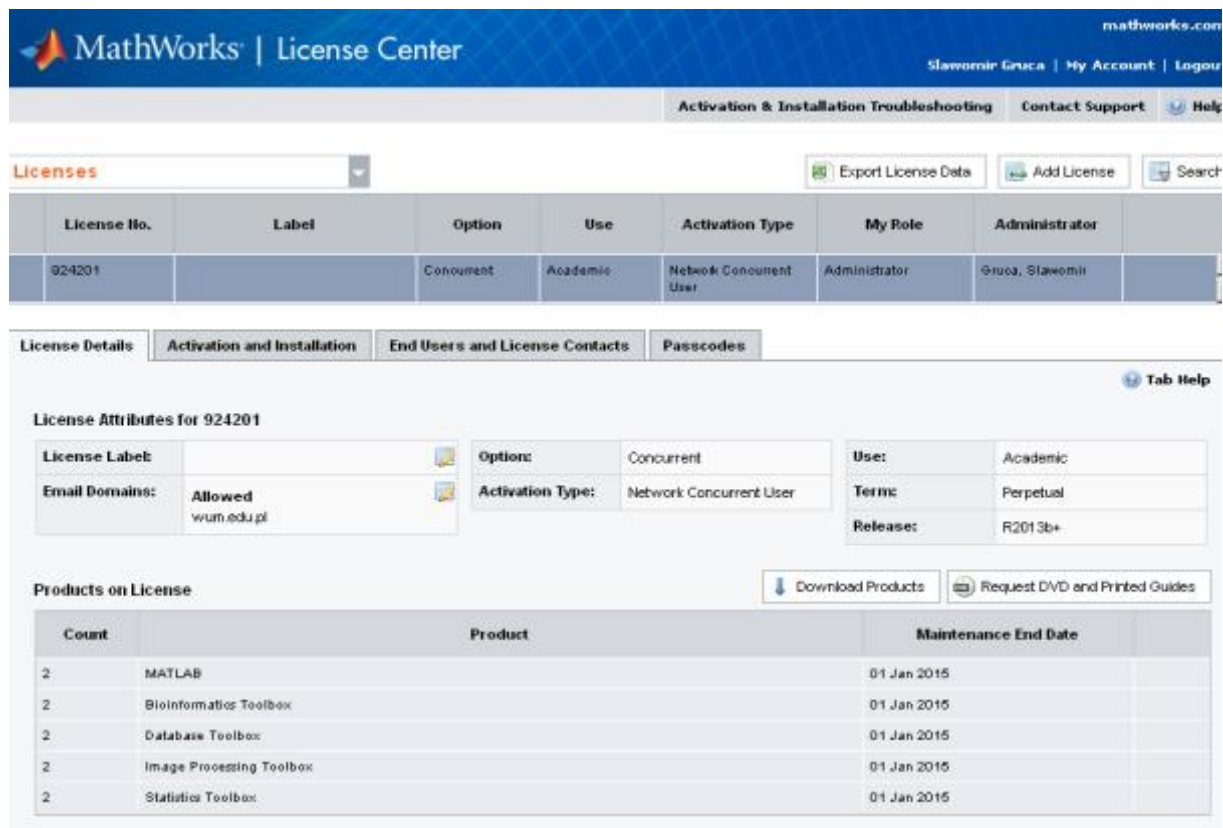


Photo 6: A screen-shot of the licence management portal of the purchased software.

3. An Uninterruptible Power Supply

Uninterruptible Power Supply is mainly designed for the equipment of Tasks T4.2 and T4.3 described in the Deliverable 4.1, but was selected in the same tender procedure as IT infrastructure.

Purchased item:

- **Uninterruptible Power Supply (UPS)** for the equipment of Tasks T4.2 and T4.3, purchased by WUM from “Delta Power Sp. z o. o.” company:
 - one UPS Delta Power GreenForce GF340 uninterruptible power supply.

The proper functioning and even lifespan of the laboratory hardware is often dependent on the quality of the electricity powering it. In order to secure the expensive laboratory equipment, namely the protein purification workstation (Task 4.2) and the automatic platform for multispectral detection of fluorescence, absorbance and luminescence on microtiter plates (Task 4.3), it was decided to purchase an advanced Uninterruptible Power Supply.

The required parameters of the UPS were selected in a way to assure high quality power conditioning and an efficient energy use. Long-life batteries were required to further minimize the environmental impact and assure operational cost savings.

The energy capacity of the batteries shall be enough to ensure the continuity of the experiments in the case of power outage. This is crucial, as the procedures sometimes last hours and involve unique biological samples.

Brief technical specification:

Parameter	Requirement(s)
class	<ul style="list-style-type: none"> VFI-SS-111
nominal power	<ul style="list-style-type: none"> at least 31 kW
efficiency for on-line mode at full power	<ul style="list-style-type: none"> at least 93%
battery lifespan	<ul style="list-style-type: none"> at least 12 years
power factor	<ul style="list-style-type: none"> at least 0.99
monitoring	<ul style="list-style-type: none"> local (LCD display) and remote (computer network)

Date of the tender announcement: 3rd January, 2014
 Date of the tender announcement decision (outcome): 24th January , 2014
 Date of signing the AEZ/365/S-299/13/42/2014 agreement: 31^s January ^t, 2014
 Remaining offers (valid ones ranked according to their price):

- “AG IT PROJECT S.C.” company – invalid offer, required technical parameter(s) not met
- “Comex spółka akcyjna” company – 75,62%
- “Alternative Technology Sp. z o. o.” company – invalid offer, missing required information
- “Spinel Sp. z o. o.” company – 52,07%
- “Fast Group Sp. z o. o.” company – 80,34%
- “Komputronik S.A.” company – 56,71%

Date of installation: 21st February, 2014
 Date of user training: 21st February, 2014

Trained persons:

No.	Name	Department
1	Sławomir Gruca	Department of Immunology, WUM
2	Marek Szlendak	Department of Power Infrastructure, WUM
3	Radosław Plewiński	IT Department, WUM

Equipment use:

The purchased equipment will complement the laboratory equipment of Task T4.2 and T4.3, and will decrease the odds of its malfunction in the case of electric line disturbances. Apart from protecting the equipment used by the researchers of the Medical University of Warsaw, it will secure the continuity of experimental procedures.

Photo 7: The Uninterruptible Power Supply for the equipment of Tasks T4.2 and T4.3 (Deliverable D4.1). This UPS provides high quality electric power for lab equipment and ensure continuity of experiments in the case of power outage.



Conclusions

We have successfully purchased and installed a modern IT infrastructure that will support our studies in the field of experimental oncology. This universal computing and storage platform will enable us to undertake ambitious studies in the area of bioinformatics, which is rapidly becoming ever more important for the medical sciences.

The delay regarding the purchase of the workstations was a result of a rarity of the SSD drives offered. At the beginning of March they had “customer sample” status by their manufacturer (Samsung) assigned, which suggests that they had not enter a regular sale phase. We were asked during the purchase phase if replacement drives could be installed in the workstations, but we decided to wait for the ones originally offered; we strongly preferred them as they would provide the performance and robustness we required for our research activities.

Having at our disposal the infrastructure, in conjunction with our expertise in data mining and analysis, algorithm development and software integration, empowers us to provide high quality custom bioinformatics support to MUW's researchers. Moreover, by providing extensive and quickly accessible mass storage space, the created platform will be used by

many researchers directly on a daily basis, making storing and moving the experimental data much more convenient.

Corresponding estimated budget

Corresponding estimated budget for acquisition of IT infrastructure

PERSONNEL, EQUIPMENT, OTHER MAJOR DIRECT COST ITEMS FOR BENEFICIARY "1" FOR 18M			
WP no	Item description	Amount/* [EUR]	Explanations
WP4	Equipment	144 592	Acquisition of modern research equipment within Task T4.7
	Other direct costs	14 445	2 air-conditioning equipment sets for computing cluster room; 4 LAN switches
	Other direct costs	5 615	Uninterruptible Power Supply
TOTAL DIRECT WP4 COST		164 652	

/ - exact costs for M1-M18 will be presented in the 1st Period Report and Form C (April 2014)*

Dr Radoslaw Zagozdzon
 New Lab Leader T4.7 Leader
 M. Sc. Slawomir Gruca IT-Professional
 M. Sc. Piotr Stawinski IT-Professional

Prof. Jakub Golab
 BASTION Project Coordinator
 Warsaw, March 2014

Attachment: Functional diagram of the BioInfo Laboratory Infrastructure

