

**Univerzita Karlova v Praze**  
**1.lékařská fakulta**

*Charles University in Prague*  
*1st Faculty of Medicine*

Studijní obor: Biomedicínská informatika  
*Study domain Biomedical informatics*

**Autoreferát dizertační práce**  
*Summary of the dissertation*



**Využití technologie GRID při zpracování medicínské  
informace**

*Utilization of GRID technology in processing of medical information*

**Mgr. Tomáš Kulháněk**

Praha 2015

**Doktorské studijní programy v biomedicině**  
*Univerzita Karlova v Praze a Akademie věd České republiky*

Obor: Biomedicínská informatika

Předseda oborové rady: Prof. RNDr. Jana Zvárová, DrSc.

Školící pracoviště: CESNET z.s.p.o., Ústav patologické fyziologie 1.LFUK

Školitel: Ing. Milan Šárek, CSc.

Konzultant: Doc. MUDr. Jiří Kofránek, CSc.

Disertační práce bude nejméně pět pracovních dnů před konáním obhajoby zveřejněna k nahlížení veřejnosti v tištěné podobě na Oddělení pro vědeckou činnost a zahraniční styky Děkanátu 1. lékařské fakulty.

# Obsah

<b>Abstrakt (česky)</b>	<b>5</b>
<b>Abstract</b>	<b>6</b>
<b>1 Introduction</b>	<b>7</b>
<b>2 Thesis Goal</b>	<b>7</b>
<b>3 Thesis Contribution</b>	<b>8</b>
<b>4 Thesis Structure</b>	<b>9</b>
<b>5 Hypothesis</b>	<b>11</b>
<b>6 Distributed computing technologies</b>	<b>11</b>
<b>7 Methods</b>	<b>12</b>
<b>8 Sharing medical information</b>	<b>12</b>
8.1 Methods to share medical images in grid . . . . .	15
<b>9 Voice Science</b>	<b>16</b>
9.1 Methods for remote analysis of human voice . . . . .	16
<b>10 Computational physiology</b>	<b>17</b>
10.1 Modeling methodology . . . . .	19
10.2 Identification of physiological systems . . . . .	20
10.3 Methods for Parameter Estimation . . . . .	21
10.3.1 Architecture of system for parameter estimation . . . .	21
10.4 Parameter Sweep . . . . .	22
<b>11 Results</b>	<b>27</b>
<b>12 Conclusion</b>	<b>28</b>
<b>13 Discussion</b>	<b>28</b>
13.0.1 Platform . . . . .	29
13.0.2 Porting . . . . .	29

<b>14 Future work</b>	<b>29</b>
<b>15 Summary</b>	<b>31</b>
<b>Reference</b>	<b>32</b>

# Abstrakt

**Název práce:** Využití technologie GRID při zpracování medicínské informace

**Autor:** Mgr. Tomáš Kulháněk

**Katedra,ústav:** Ústav patologické fyziologie 1.LFUK

**Školitel:** Ing. Milan Šárek, CSc., CESNET z.s.p.o.

**Konzultant:** Doc. MUDr. Jiří Kofránek, CSc.

## Abstrakt:

Práce prezentuje výzkum a výsledky využití technologií, které umožňují sdílet výpočetní a úložné kapacity a to v oblasti biomedicínského výzkumu. Vedle technologie GRID se rozvíjí virtualizačních technologie (VMWare, XEN, ...), které dodali distribuovaným systémům novou vlastnost zdání vlastnictví a přímé kontroly konfigurovatelné infrastruktury jako služby, jež se dnes shrnují pod společný pojem CLOUD computing. V práci jsou diskutovány teoretické limity distribuovaných systémů a paralelních výpočtů v nich tak i praktické výsledky ve vybraných oblastech. V oblasti výměny medicínských snímků a souvisejících zdravotních záznamů byla ukázána snadná integrovatelnost se stávajícími systémy při respektování požadavků na bezpečnost dat. V oblasti analýzy a zpracování lidského hlasu v reálném čase byla ukázána možnost poskytování nadstandardních výpočetních služeb pro komunitu uživatelů. V oblasti modelování fyziologických systémů je prezentován systém pro odhad parametrů a identifikaci fyziologických systémů komplexních modelů, které by byli obtížně řešitelné za použití klasických metod.

## Klíčová slova:

grid-computing, cloud-computing, matematické modelování, výpočetní fyziologie, fonetogram

# Abstract

**Title:** Utilization of GRID technology in processing medical information

**Author:** Tomáš Kulháněk

**Department:** Institute of pathological Physiology

**Supervisor:** Ing. Milan Šárek, CSc., CESNET z.s.p.o

**Consultant:** Doc. MUDr. Jiří Kofránek, CSc.

## Abstract:

Two work presents results, which are researched in the field of utilization technology who offers distributed computing and storage capacity to exchange medical information, demanding computation and data exchange. Next to the GRID technology. The virtualization technology evolved and were utilized more massively next to the GRID technology and these gives distributed systems a new quality and services called today with the name CLOUD. This work summarizes result of different projects which implements selected technologies in the field of GRID and CLOUD to systems which are used in medical education and research and in neighboring fields. The multidisciplinary projects were solved within the cooperation between association CESNET, First Faculty of Medicine of Charles University in Prague and Academy of Performing Arts in Prague.

**Keywords:** grid, cloud, computational physiology, phonetogram

# 1 Introduction

The *grid-computing* is usually defined as sharing computational and data storage resources across organizational boundaries. In recent years, the development of virtualization technologies enhances the availability of services provided by grid-computing and additionally enabled an evolution of so called *cloud-computing*, which can utilize virtual environment on real powerful computing infrastructure too. Based on the development of technologies and also philosophy of providing them to end users, this thesis focus on multidisciplinary research related to grid-computing as well as to cloud-computing and it's utilization in biomedical research and application related to processing of medical information.

The term "medical information" is too wide and further work focuses on the following selected areas, which were part of: (1) exchange and processing of medical images, (2) analysis of human voice and (3) modeling and simulation of human physiology.

The author's work was published in a series of peer-reviewed papers of international journals and peer-reviewed conference proceedings [1, 2, 3, 4, 5, 6, 7] which are attached into this work as appendices. The author's work and contribution was also presented in international conferences and published in the respective proceedings and transactions [8, 9, 10, 11]. The work was also popularized on the local and regional conferences and their respective proceedings [12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23]. Author contributed to the utility model registered by the Czech Industrial Property Office [24].

# 2 Thesis Goal

The hypothesis stated by this thesis is that the technologies related to grid-computing and cloud-computing may improve processing of medical information to perform demanding tasks which are almost impossible or may need onerous effort to achieve using classical local or institutional resources.

The aim of the thesis is to discuss the hypothesis in different areas of biomedical research and it's application. The thesis tries to find answers to the following questions:

- *What are the possible benefits and limitation of utilizing grid-computing (or cloud-computing) technology for processing medical information?* In the time of starting the work on this thesis, it was believed that grid-

computing may be an answer to scalability issues of exchanging large amount of data or doing demanding long-term computation.

- *What additional service should provide grid-computing or cloud-computing infrastructure?* In the time of starting the work on this thesis, the grid-computing infrastructure was used mainly to store, exchange and process medical data, mainly images in DICOM format and protocol (details in section 8).
- *How should be designed the architecture of such services?* The main type of service design was service oriented architecture (SOA), while a focus was again get to plain old data objects represented e.g. in JSON format and processed by RESTful web services following REST architectural constraints (explained more in section ??).

### 3 Thesis Contribution

The author claims that the following contribution was made to the state of the art of biomedical informatics related to grid-computing and cloud-computing.

- Proposal of grid infrastructure and pilot implementation of grid-based system of exchanging medical images integrated with existing distributed systems. The results were published as [1] and popularized as [12, 13, 14]. The author of this thesis customized the existing results of Globus MEDICUS project and deployed a grid application in the servers networked via academic network CESNET and integrated with existing regional PACS system.
- Pilot implementation of more generic infrastructure as a service for the community within the biomedical research [3, 19]. The author of this thesis proposed the idea to share the physical resources and to provide virtual environment for specific needs of particular use-cases. With development of virtualization techniques and cloud-computing technologies the pilot infrastructure were tested on examples of selected research projects.
- Proposal of software architecture and implementation of web-based service for real-time remote analysis of human voice. The results were published as [2] and popularized as [16, 20]. The author of this thesis contributed to the idea of utilizing interactive remote access to the



computing environment. For this reason it was identified requirements and customized the existing network protocol to transfer voice signal losslessly. Other co-authors implemented the algorithms and application to analyse voice signal.

- Improved methodology for modeling of complex physiological systems [5, 6, 11, 10]. Author of this thesis contributed to the idea of building complex mathematical models from the basic components and keep them in an understandable and maintainable form utilizing some known good practices (patterns) and prevent bad practices (antipatterns) known from object-oriented programming. Author advised and implemented several basic blocks and models of pulsatile cardiovascular system in Modelica language. The other co-authors implemented the library to model physiology using integrative approach and implemented the complex models integrating different domains together.
- Design and implementation of system to estimate parameters of complex mathematical models to validate or calibrate models of human physiology published as [4] and gradual development of related technologies were published and popularized as [8, 9, 18, 22]. Author of this thesis designed the architecture for distributed parameter estimation algorithm and implemented pilot deployment utilizing scientific cloud-computing infrastructure and integrated complex model simulation with numerical identification algorithms. Other co-authors implemented complex models of human physiology in Modelica language and tested several algorithms for parameter estimation.
- Improved mathematical model of oxygen, carbon dioxide and hydrogen ion binding to Hemoglobin [7]. Author of this thesis implemented this model in Modelica and calibrated the parameters of the model. Other co-authors analysed and proposed new mathematical model based on basic physical and chemical laws and relation published in literature.
- Simulation of complex models of human physiology as part of virtual simulator on portable and mobile devices utilizing cloud-computing [9, 21]. Author of this thesis contributed to the idea of hybrid architecture of web simulators - utilizing the infrastructure for parameter estimation to simulate complex model remotely and process/visualize the results locally. Other co-authors implemented complex models of

human physiology and specified and implemented simulation scenarios for educational purposes.

- Virtual patient simulator prototype registered as utility model by the Industrial Property Office in the Czech Republic [24]. Author contributed to the development of prototype virtual simulator within module of cooperation of multiple instances within virtual classroom. Other co-authors designed and implemented models of human physiology, clinically relevant educational scenarios and implemented 3D visualization of selected scenarios using game engine Unity 3D<sup>1</sup>.

## 4 Thesis Structure

This thesis is interdisciplinary, therefore the following chapters will cover the topics not-only from technical and computer-science point of view, but touches some topics related to the medical science mainly human physiology and it's mathematical models and simulations. The chapter ?? provides an overview of the state of the art in the theory of computation, parallel computation, distributed computation and focus on task-parallelization in grid-computing and cloud-computing infrastructure.

Introduction, particular methods and results to selected areas of biomedical application and research are introduced in chapter 8 for sharing medical images, in chapter 9 about voice science and chapter 10 about computational physiology.

The chapter 11 summarizes general results obtained by the research methods in specific areas of biomedical research and applications. The chapter 12 discuss achievements and answers hypothesis and questions stated at the beginning of the work and recommends further direction of the research effort.

The appendices contain the selected papers [1, 2, 3, 4, 5, 6, 7] which are most relevant to the topic of this thesis and which were published in international peer-reviewed journals or in peer-reviewed conference proceedings:

**Appendix ??** is the paper [1] *Processing of Medical Images in Virtual Distributed Environment* published by ACM as part of the proceedings of the 2009 Euro American Conference on Telematics and Information Systems: New Opportunities to increase Digital Citizenship.

---

<sup>1</sup><http://unity3d.com/> accessed March 2015

**Appendix ??** is the paper [2] *Remote Analysis of Human Voice – Lossless Sound Recording Redirection* published in Analysis of Biomedical Signals and Images. Proceedings of 20th International EURASIP Conference (BIOSIGNAL).

**Appendix 15** is the paper [3] *Infrastructure for data storage and computation in biomedical research* published by Euromise s.r.o. in the European Journal of Biomedical Informatics.

**Appendix ??** is the paper [4] *Parameter estimation of complex mathematical models of human physiology using remote simulation distributed in scientific cloud* published in the IEEE Xplore Digital Library as part of the proceedings of the 2014 IEEE-EMBS International Conference on Biomedical and Health Informatics.

**Appendix ??** is the paper [5] *Modeling of short-term mechanism of arterial pressure control in the cardiovascular system: Object-oriented and acausal approach* published by ELSEVIER in Computers in Biology and Medicine 2014, **IF(2013): 1.475**.

**Appendix ??** is the paper [6] *Simple models of the cardiovascular system for educational and research purposes* published in Mefanet Journal 2014.

**Appendix ??** is the paper [7] *Adair-Based Hemoglobin Equilibrium with Oxygen, Carbon Dioxide and Hydrogen Ion Activity* published in Scandinavian Journal of Clinical and Laboratory Investigation 2014, **IF(2013): 2.009**.

## 5 Hypothesis

To summarize this section. If there will be technological speedup, this will impact mainly the class of problems which are solvable by polynomial algorithm. For the problems where the computation needs tremendous amount of time, because current known algorithm is exponential, there can be used non-exact methods to find at least some solution if not the exact one.

- The *heuristic methods* tries to eliminate the number of steps of computation by some implicit or explicit knowledge of the specific problem itself E.g. eliminating solution classes that seems not to go to optimal solution. With combination of brute-search the heuristic methods reduce the size of all possible solution candidates to check.
- The *randomization methods* use non-deterministic methods in some level of computation. E.g. Monte-Carlo method is used to compute problems using pseudo-random generated values and after several iterations statistical methods are used to compute expected value and standard deviation.
- *Restriction on input data* - is another form of using the explicit knowledge of the problem instance and it may reduce all possible values to be checked.
- *Approximation algorithm* - may find not only some good solution, but can quantify how far from the optimal solution the found is good with some degree of probability.

To summarize this section; parallel computing can introduce speedup on current computational technology and some computation problems may become feasible. Also overhead caused by parallelization and fraction of non-parallelizable parts should be considered as it may degrade expected speedup. In case of exponential algorithm (e.g. for NP-complete problems) the speedup will increase the size of solvable problem only slightly (see table ??) and some problems cannot be (or it is believed) significantly speeded up by parallel computing. In further text a focus will be given mainly to task parallelism and distributed computing.

## **6 Distributed computing technologies**

Distributed computing is based on the idea to spread the computation task into set of computers which are connected via computational network.

## 7 Methods

The general issue of utilizing grid computing or cloud computing infrastructure is to select appropriate method to integrate domain specific computation into the grid or cloud infrastructure of a concrete provider.

A lot of tools are already available within current grid infrastructure including open-source or licensed software for computation. A list of available application is usually given by the local scientific provider<sup>2</sup> or application database are available in broader environment e.g. in EGI.eu application database<sup>3</sup>. Additionally workflow systems and scientific gateways mentioned in section ?? tries to hide the complexity of grid-computing or cloud-computing infrastructure and may be used to integrate specific domain too. The programming model of parallel computing and/or distributed computing (in section ?? and ??) needs to be followed when designing new application utilizing benefits of grid-computing and/or cloud computing.

The general approach to port application to grid infrastructure is to automatize what can be automatized, i.e. make scripts, configure system, prepare some UI, integrate with existing applications, utilize protocol compatibility etc. An effort to obtain first results is high, however for further computational request, the prepared templates, scripts are reused and effort is much lower.

## 8 Sharing medical information

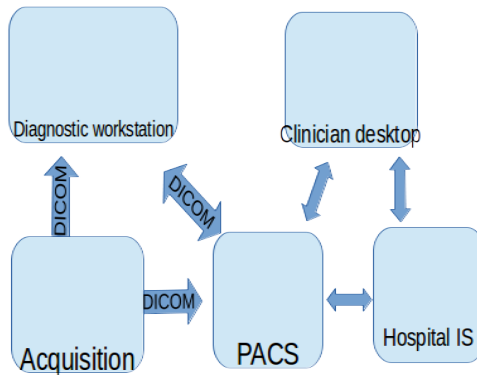
Use cases related to digital medical images involves the image acquisition, preprocessing, storing and searching. Clinicians use patient image mainly for visualization and diagnostic purposes. Computer assisted methods facilitate the diagnostic process and involves image enhancement (to reduce image noise and increases the contrast), image segmentation (to separate different types of structures from background and from each other), quantification methods (to determines the structure shape, size, volume), registration methods (to process and join multiple different images into one). Comprehensive concepts and digital techniques in medical imaging are presented e.g. in book edited by I.N.Bankman[25].

---

<sup>2</sup>applications available in CESNET METACENTRUM <https://wiki.metacentrum.cz/wiki/Kategorie:Applications> accessed February 2015

<sup>3</sup><https://appdb.egi.eu/> accessed February 2015

Acquisition of the medical image is covered with different modalities (different types of equipment and sensors) by radiologists or other specialists. DICOM<sup>4</sup> format and protocol becomes the de-facto an industrial standard to exchange medical images electronically and picture archiving communication systems (PACS) holding the acquired DICOM images with metadata and description noted by experts are currently part of information systems in hospitals. See the typical workflow of medical image in hospital in fig. 1.



Obrázek 1: Typical workflow of medical image in hospital. Data acquisition is made by modalities (magnetic resonance, ultrasonography, X-ray radiography, etc.) and using DICOM format and protocol it can be directly transferred and visualized by diagnostic workstation. With metadata filled by an expert physician the image is stored in PACS. Other desktops within hospital can retrieve the image and review the report. The hospital information system may be involved in other workflows and communicate with other formats and standards (HL7,...)

As the data processed in hospital information systems contains sensitive information of real patients, these are protected and processing and storing

<sup>4</sup>DICOM: <http://dicom.nema.org/> accessed January 2015

is regulated by the national or international laws or agreements. Development of telecommunication and network technologies enabled telemedicine – providing healthcare over remote distance. It requires to share and exchange sensitive data of real patient among different healthcare providers and additionally such data may be very valuable for further research. Security and encryption should be addressed and DICOM standard itself doesn't solve security issues appropriately, thus encryption during transferring the data over computer network must be ensured by other techniques.

In the Czech Republic, there exists several projects in production interconnecting different hospitals, clinics and other healthcare organization to exchange medical images. Project ePACS allows interconnecting each participant's PACS system via dedicated VPN channel to the central node and exchange of medical images are realized by routing the data flow from one VPN channel to the other<sup>5</sup>. Another approach is used in the project MEDIMED held by Masaryk University in Brno. Instead of dedicated VPN channel, they use SSL encryption over standard TCP/IP communication and regional hospitals and healthcare providers are interconnected via the MEDIMED servers [26]. In other countries, there were tested cross-border teleradiology in projects Baltic e-health, R-Bay and others [27, 28]. These projects are focused on sharing the medical images and other knowledge and information.

Access to the wide range of medical images is needed for research of new processing and diagnostic methods, rare diseases, developing new detection algorithm etc. DICOM records "de-identified" (identification of patient records are deleted, only date of the birth and other data are kept) or anonymized (additional information are manipulated to prevent disclosure) for research purposes to protect sensitive personal data, but keep important information for research purposes. The Globus MEDICUS project published by Erberich et al. [29, 30] is based on Globus Toolkit middleware to federate clinical and research application via a grid-computing infrastructure. Currently the project is hibernated since 2008 and no further development was published<sup>6</sup>. Similar effort was done with a project Medical Data Manager which uses gLite grid middleware published by Duque, Montagnat et al. [31, 32]<sup>7</sup> or MediGRID project published by Krefting et al. [33, 34]. Additionally to the sharing medical images, processing of images within selected use-cases supported by the grid-computing infrastructure is introduced [34]. Health-e-child project aimed to interconnect research institution and hospi-

---

<sup>5</sup>ePACS: <http://www.epacs.cz>, accessed January 2015

<sup>6</sup><https://dev.globus.org/wiki/Incubator/MEDICUS> accessed February 2015

<sup>7</sup><http://modalis.i3s.unice.fr/softwares/mdm/start> accessed February 2015



tals in UK, France and Italy for the purpose of grid-based healthcare platform for pediatric health-care [35]. Neurist project developed architecture and connects clinicians and researchers to improve research and treating of cerebral aneurysm to provide tools to analyze and interpret patient data and researcher can have access to set of aneurysm data, published by Benkner et al.[36]. SEAGRIN research project aimed to share knowledge mainly for educational purposes in semi-formally described semantics and such proposal and implementation was published by Kuba et al.[37].

Storing the sensitive medical information even de-identified or anonymised is usually restricted and this lead to an idea to store such information within trusted institution e.g. hospital and move and facilitate deployment of the grid services storing medical data to that institution. E.g. pre-installed virtual machines can contain grid-services and deployed in as a sealed grid as proposed by Kuba et al. [38].

To summarize this section, digital medical image acquisition, store, exchange and processing became common in the past years and is currently using distributed computing techniques. There are several efforts to implement medical data management within grid or cloud infrastructure for research purposes and integrate them with the production infrastructures. Security is solved by authentication, authorization mechanism as well as by encrypting the data and/or de-identification or anonymization but keeping minimal information required for research purpose. A related question is how easily the previously mentioned grid-based technologies can be integrated with current systems in hospitals or institutions. The following section describes selected methods used to integrate a pilot deployment of Globus MEDICUS with current regional system for exchanging medical images - MEDIMED.

## 8.1 Methods to share medical images in grid

Globus toolkit belongs to the group of most used grid middleware (see section ??). The core services included in Globus Toolkit are GridFTP – grid extension to file transfer protocol(FTP) implements strategies such as *stripping data* into multiple pieces, *parallel transfer of data* utilizing stripped data parts to be transfered via different channels, *partial file transfer* some application may not need to access the whole file, but a smaller portion of it. etc. [39, 40]. Other core services are Replica Location Service aiming to localize data, Globus Resource Allocation Management (GRAM) provides web service and proxies to the lower level job schedulers implementation [39].

Next to core services, the domain specific services might be implemented for the purpose of application using the open grid service architecture (OGSA). Globus MEDICUS [29, 30] contains a DICOM Grid Interface Service (DGIS) and integrates the open source PixelMed™ Java DICOM Toolkit<sup>8</sup> into a web service communicating via DICOM protocol and on the other side it forwards the queries to underlying services within Globus toolkit.

DGIS behaves as a gateway to a grid infrastructure. Because communication via DICOM protocol is not secured, the DGIS is recommended to be installed on the location of the PACS system or DICOM ready modality or software. When a DICOM study is uploaded into DGIS, it is anonymized and stored and a record is made into another services Meta Catalog service which resides in the same domain or anywhere in grid accessible via Globus Toolkit. Such anonymized database of DICOM records can be used to query via DGIS interface and to e.g. integrate with web based application showing records for research purposes, authentication and authorization can be done in this level. To integrate this system with existing system for sharing the medical images (e.g. the MediMed project[26]) the special client software "RediMed console" needs to be installed next to the DGIS and configured it as a local PACS system whose records might be exchanged to other MediMed participants. The results of this particular deployment and integration is in section ??.

## 9 Voice Science

With introduction of objective data analysis and laryngoscopy methods the voice science emphasized the cooperation among laryngologist, speech pathologist and voice teacher. The human voice ranges from 50 Hz to something about 1000 Hz, but there are large individual variation. For analysis of digitally recorded voice, either habitual or singing, the Discrete Fourier Transformation(DFT) is used to produce frequency and amplitude analysis of recorded input voice samples. One of the most used class algorithm to compute DFT is class of Fast Fourier Transformation with computational complexity  $O(n \log(n))$  [41, 42]. The result of analysis can be visualized in a voice range profile and there can be seen significant difference between untrained and trained voice as well as quantitatively seen some disorders [43, 44].

Other methods to analyse vocal chords is laryngoscopy. The videostroboscopy and high speed video in laryngoscope methods produce video

<sup>8</sup><http://www.pixelmed.com/> accessed February 2015

for analysing the real movement of vocal chords. The videokymography method introduced by Švec et al. complements the videostroboscopy and allows to visualize and analyze movement of vocal cords recorded by high speed camera on standard TV or monitor with an artificial image built from recorded sequence of selected section [45, 46].

In case of recorded sound and further analysis there is a question about how such a service can be integrated in grid-computing or cloud-computing environment to provide access to a complex application for non-technical voice specialists. Additionally, the analytical software was already developed and calibrated for selected sorts of microphones in MS Windows platform [47, 48]. Therefore I proposed and implemented a method that provides access to the analytical software remotely. The section 9.1 describes how the analytical software was customized with a remote desktop protocol (RDP). Results are described in section ?? . Similar approach might be used for processing the video recordings from laryngoscope, however, the practical limits are discussed in section 12.

## 9.1 Methods for remote analysis of human voice

Terminal access to some remote computational capabilities, e.g. remote command-line or remote execution is another integration strategy to some remote infrastructure. Secure Shell (SSH) is used to establish secure channel via unsecured network (e.g. the Internet) from SSH client to SSH server and it is basic method to access grid-computing infrastructure.

Remote Desktop Protocol(RDP) is a proprietary protocol for desktop sharing developed primarily in Microsoft Windows platform, however, today clients and servers exists for several other platforms. Next to remote command-line, remote execution it allows to access remote graphical desktop environment. The software for parameterized Voice Range Profile (ParVRP) and Voice Range Profile in Real time (RealVoiceLab) was already developed and calibrated for selected sorts of microphones in MS Windows platform by Fric et al.[47, 48]. The implementation is done in MATLAB environment utilizing Signal Processing Toolbox<sup>9</sup> and compiled with MATLAB Compiler and distributed as an executable.

Instead of migrating the application into some compatible platform for grid-middleware, a virtual machine was introduced and access to the software is provided via RDP protocol. RDP itself contains redirection of several services, e.g. sound recording or drive access. Because the default sound recording redirection introduces some sound degradation without control, I

<sup>9</sup><http://www.mathworks.com/products/signal/> accessed February 2015

proposed, implemented and integrated the custom RDP plugin with the ParVRP and RealVoiceLab software to redirect the sound recording without loss of information. Technical details are in Appendix ??.

The computation of frequencies and amplitude from the recorded samples utilizes Fast Fourier Transformation which has time complexity  $O(n \log(n))$  and current implementation are fast on any even modile devices. The benefit from deploying such application in distributed infrastructure is immediate access to updated software and a collection of anonymized records of voice samples with analyzed results for further research and education purposes.

This type of application can be packaged as virtual machine template and configured within different types of cloud infrastructures and together with a script or web portal the on-demand deployment can be automated. The client part (RDP client) needs to connect to the appropriate instance. The results of such deployment are discussed in section ??.

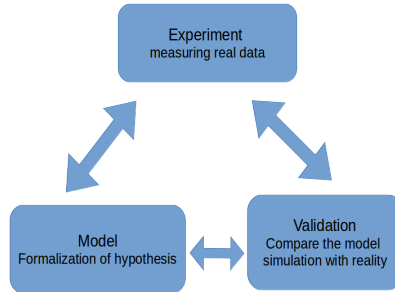
## 10 Computational physiology

A mathematical formalization of the fundamental knowledge and relation among biological system - mathematical model - is used as a base abstraction to utilize current discoveries of the genomics and proteomics and formalize the knowledge and construct a "Physiome Model". Model by it's definition is simplification of the complex reality.

Constructing the models and integrating them into complex entity which can be used for further purposes is schematically illustrated in fig. 2. The measurements are done in laboratories or in hospitals. Lumped parameter models are usually represented as ordinary differential equations and differential algebraic equations and characterize the reality as topology of discrete elements. The imaging methods for processing and analysis (section 8) are used to construct 3D models from segmentation and generating of mesh representation connected to physical principles.

Application of the mathematical modelling techniques towards the biomedical research is sometimes called as systems biology approach combining the reductionism and integration as denoted by Kohl et al.[49]. Application towards the clinical praxis include the quantification of the diagnostic index or treatment strategy and it is a goal to develop tools, database models and methods of several Physiome projects, e.g. VPH-Physiome project presented by Hunter et al.[50].

One of the earliest complex and integrative modelling effort was a model of circulation and it's regulation published by Guyton et al. in 1972



Obrázek 2: Schematic illustration of scientific process. The experiments produces data which are interpreted and hypothesis is formalized as a model. Validation compares the model simulation with experiment, if model satisfies the criteria - is in agreement with real experiments, then the validated model can be used for other purposes.

[51] which via derivative and technological upgrade continues as "Human Model" or "HumMod" introduced by Hester et al. [52, 53] with a focus on integration effort. Different approach of modelling human physiology is a database of smaller models focusing on some particular physiological phenomenon. E.g. the NSR Physiome project introduces JSIM<sup>10</sup> Java based simulation system to support modeling in physiology. Repository of several hundred of models were published using this system [54]. The similar effort is done by IUPS Physiome project and repository of models are based on XML standard languages CellML and FieldML [55, 56]. The Systems Biology Markup Language (SBML) is used for modeling biological system at the level of biochemical reaction and regulatory network and another database collects several hundreds of curated and non-curated models [57, 58].

JSIM, CellML, SBML or HumMod are domain specific languages and the tools able to work with them are primarily developed within physiological or systems biological communities. Other authors use commercial or industry standard tools for mathematical modelling and computing. E.g. Kofranek et al. describes Guyton's 1972 model in MATLAB<sup>®</sup> Simulink [59] and the derivative HumMod in acausal object-oriented Modelica language [60, 10]. Fernandez et al. describes models of cardiovascular pulsatile system

<sup>10</sup>JSIM: <http://www.physiome.org/jsim/> accessed January 2015

using MATLAB Simscape [61] and recently in Modelica [62].

Thus there is an open debate whether in-house domain specific language and tools like JSIM, CellML and FieldML, SBML or HumMod reached it's capabilities for representing complex models. Only the HumMod reached the integrative approach building the complex integrative model of human physiology using lumped parameter approach. I contributed to the idea of key features which involves acausal modeling technique and object orientation which keeps the complex model structure decomposed into understandable and maintainable parts and allows to cover complexity of models like HumMod.

The methods and examples of modeling cardiovascular system are described in the next section 10.1. The methods of estimating parameters of complex models are described in section 10.3 and particular results are described in section ??.

## 10.1 Modeling methodology

The methodology of formalizing mathematical models is influenced by the abilities of underlying modeling language used. The Modelica language is an object-oriented, equation based and acausal modeling language standardized by Modelica association<sup>11</sup>.

The paper [5] *Modelling of Short-term Mechanism of arterial pressure in the cardiovascular system: Object-oriented and acausal approach* in Appendix ?? published disputation about causal and acausal approach in using Modelica for modeling pulsatile cardiovascular (CVS) system.

The paper [6] *Simple Models of the Cardiovascular System for Educational and Research Purposes* in Appendix ?? published detailed methodology of modeling pulsatile CVS in Modelica.

Common guide to the Modelica language and it's capabilities are in the book of Fritzson [63] or in the on-line book by M.Tiller [64].

## 10.2 Identification of physiological systems

Usually some knowledge of the system - the structure is available and unknown coefficients (parameters) remain unknown. Once the model is formalized and constructed, further problem is to estimate the model parameters so that the model reproduces real world system. This procedure is sometimes called system identification and the objective of parameter estimation

---

<sup>11</sup><http://www.modelica.org> accessed February 2015

is usually to minimize the following function (to find least squares of the differences between predicted and measured values):

$$f(\vec{p}) = \sum_{i=1}^n (M(t_i, \vec{p}) - d(t_i))^2 \rightarrow \min \quad (1)$$

where  $\vec{p}$  is vector of values of parameters,  $M(t_i, \vec{p})$  is model simulated at time  $t_i$  with the given parameter values  $\vec{p}$  and  $d(t_i)$  is the measured experimental value at time  $t_i$ . In general mathematical models of biological systems are in most cases non-linear and can be non-differentiable thus local optimization methods. Algorithmically, this problem was shown to belonging to the *NP-complete* problems [65] thus the best known exact algorithm is based on brute force search - e.g. trying all possible values of parameters and simulate the model with them and find minimum of the objective function 1. Further reading about parameter estimation and system identification can be found in book edited by Eykhoff [66] or book of Khoo [67, p. 159].

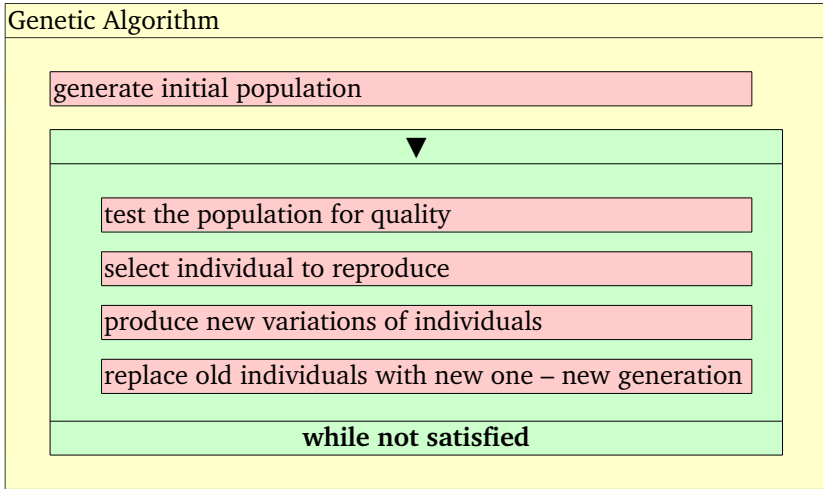
, therefore global optimization methods should be used and are based on some heuristic to reduce the number of simulation. The evolution strategies were identified as robust with potential to utilize parallel computing as shown by Moles et al. [68]

After the parameter estimation a further problems arise with structural identifiability and analysis of sensitivity to the estimated parameter values [67, p. 176].

Parameter estimation and further analysis methods are part of specialized mathematical software. E.g. Pruet et al. used Metropolis algorithm to produce a distribution of parameters to calibrate the model of human cardiovascular physiology, which were further tested against predictive ability of circulatory failure and statistical methods performed in the software Wolfram *Mathematica* [69]. The iterative improvement method in the software MATLAB Simulink® was used in estimating 2 parameters of simple cardiovascular model by Takahashi et al. [70]. Several methods were compared in estimating multiple parameters of cardiovascular system in MATLAB Simulink® by Abbass et al. [71].

Maffioletti et al. published GC3Pie framework utilizing evolutionary algorithms and introduced workflow to identify parameters of models for economical predictions using grid computing [72]. Humphrey et al. calibrated hydrology models utilizing commercial Windows Azure cloud computing infrastructure with a significant speedup on the modified dynamically dimensioned search algorithm [73, 74].

### 10.3 Methods for Parameter Estimation



Obrázek 3: Kopenogram of genetic algorithm.

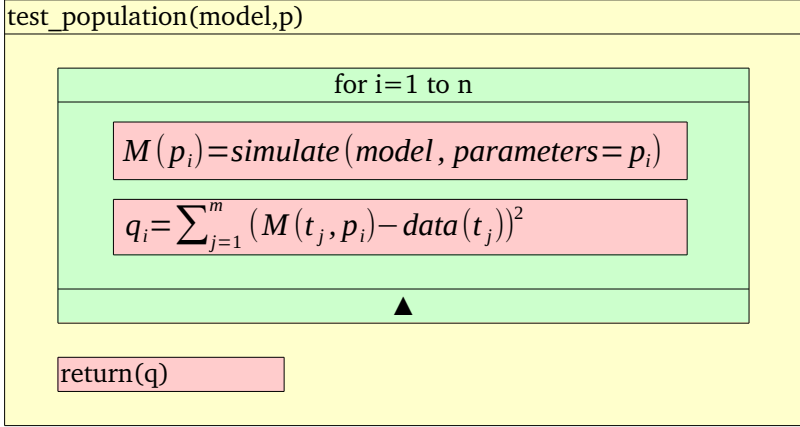
Evolutionary algorithms can be used as heuristic strategy for finding global minimum or maximum and it can be used to estimate the parameters of the model. Genetic algorithm a type of evolutionary algorithm which encodes individuals as binary string was introduced e.g. by Holland[75] and the algorithm steps are schematically presented in figure 3.

The iteration within the loop "▼... *while not satisfied*" depends on previous iteration, thus it cannot be parallelized. The step *test the population for quality* has algorithmical structure in fig.4 for parameter estimation tasks. Each iteration in the loop "*for i=1 to n*" is independent and a therefore loop parallelism (section ??) can be utilized and implemented here.

#### 10.3.1 Architecture of system for parameter estimation

Proposed architecture of the system for parameter estimation (fig. 5) was influenced by the need of some interactivity and overall accessibility for users which is fulfilled by the web UI. The key part of the system in opposite side is a model exported into a binary platform dependent library. The specific model of a studied system implemented in Modelica is exported into standard Functional Mockup Unit(FMU) with is standardized XML metadata





Obrázek 4: Kopenogram of genetic algorithm and specific test of the population for quality in case of parameter estimation. Model is simulated according to individual  $i$  with parameters  $p_i$  and the quality  $q_i$  is counted per the objective function 1.

packaged together with binary library .DLL (or .SO) following standardized API [76]<sup>12</sup>. In the time of writing the thesis the most stable Modelica tool export was Dymola<sup>13</sup> export to FMU for MS Windows platform.

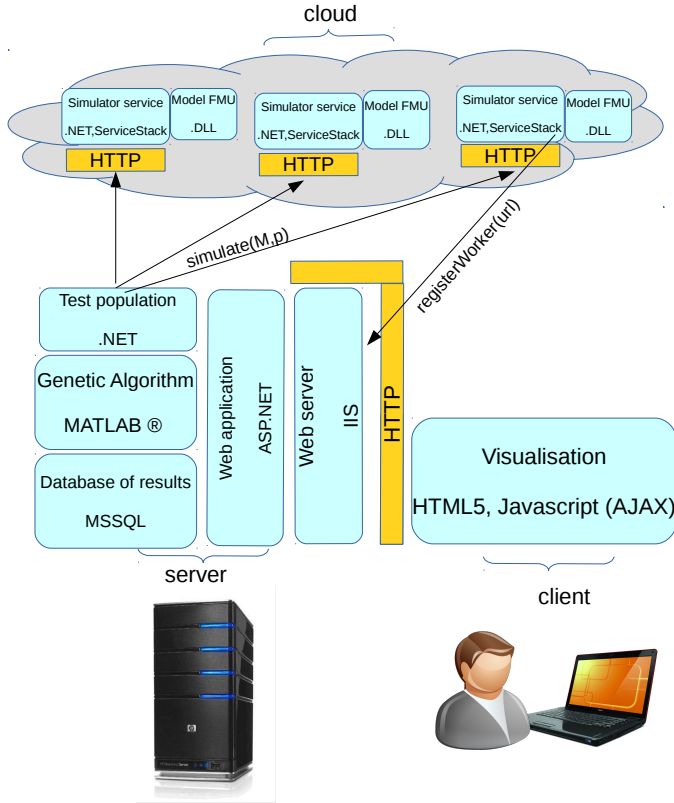
The parallelization is implemented using threads in *test\_population* method which within a loop follows fork/join pattern – the created threads simultaneously asks for simulation results with a parameter set and main process waits until all results are returned to compute full vector of quality evaluation  $q$ .

Packaged with .NET ServiceStack framework<sup>14</sup> it exposes a simulation functionality as a RESTful web service which can be accessed and orchestrated by the *test\_population* algorithm. The implementation of genetic algorithm is reused from MATLAB<sup>TM</sup> and with a database of results in a SQL database is integrated with ASP.NET web application presenting a web user interface and functionality to a user. The result of applying the methods and deploying the designed system in local cluster and cloud computing infrastructure is described in section ??.

<sup>12</sup><https://www.fmi-standard.org/> accessed February 2015

<sup>13</sup><http://www.dynasim.se> - Dymola tool, accessed March 2015

<sup>14</sup><https://servicestack.net/> accessed February 2015

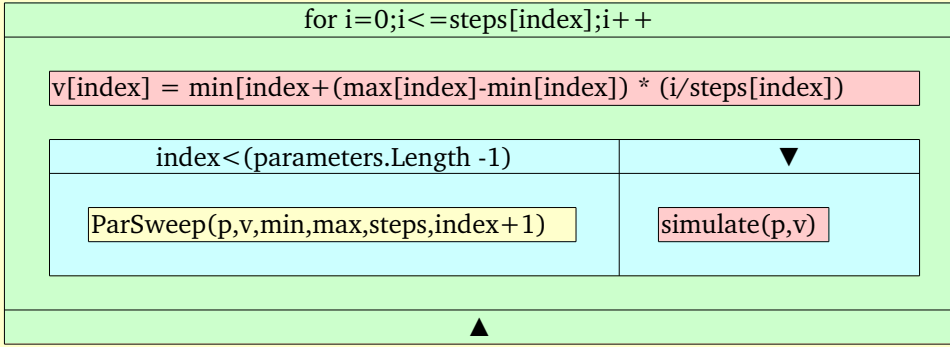


Obrázek 5: Architecture of the system employing genetic algorithm and distributing the *simulate* task into cloud computing environment.

## 10.4 Parameter Sweep

Parameter sweep (PS) is one of the techniques used for sensitivity and uncertainty analysis which is based on changing selected parameters and simulating whole model and quantifying the change on model behavior with different parameters. Uncertainty and sensitivity analysis tries to determine how a change of the value of parameter will contribute to the model output and how the estimation of parameter values are robust to errors of measurement of the real data. Various methods to do uncertainty and sensitivity analysis can be found e.g. in a reviews by Helton et al. [77] or a books by Saltelli et al. [78, 79].

ParSweep(p,v,min,max,steps,index)



Obrázek 6: Kopenogram of recursive parameter sweep algorithm.  $p, v, \min, \max$  and  $\text{steps}$  are arrays with the same dimension holding parameter name, value, starting and stopping value and number of steps which needs to be performed between starting and stopping value per each index.

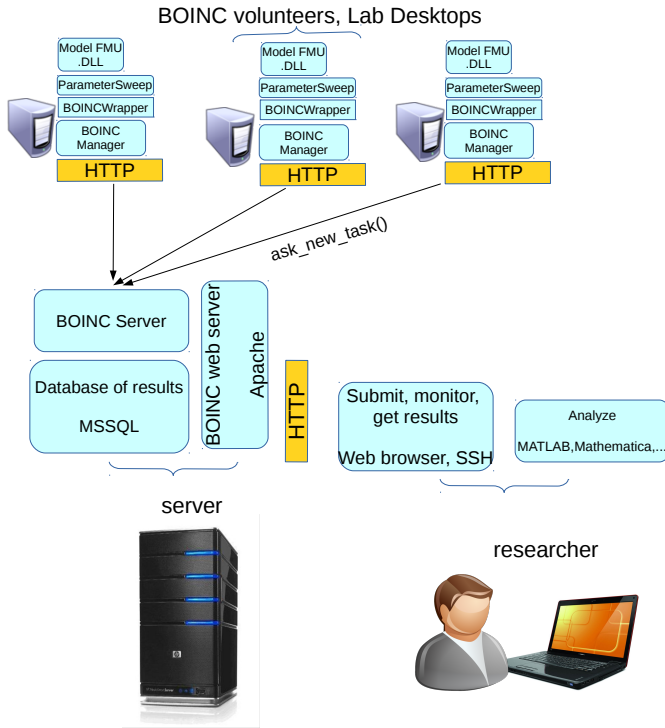
Recursive algorithm of parameter sweep for exploring parameter space (in fig.6) generates tremendous number of simulation. Presuming that *simulate* operation takes constant time for any parameters (which is not true in general) the time complexity of PS is  $O(\prod_{i=1}^n \text{steps}_i) \approx O(k^n)$  where  $k = \max_{i=1}^n (\text{steps}_i)$  and  $n$  is number of parameters to be swept. E.g. for 1000 values for each parameter:  $O(1000^n)$ . The large number of distinct simulation can take tremendous time on single computer. However, in contrast to parameter estimation, each of the simulation is independent and PS algorithm is determined as embarrassingly parallel and is implemented in many grid-computing projects and workflows e.g. P-Grade portal as published by Kacsuk et al.[80].

To perform parameter sweep algorithm on the models of human physiology in Modelica language an export from the Modelica is needed. The FMU standard supported by many tools exports FMU as

a BOINC platform[81]<sup>15</sup> is customized following the task parallelism and master/worker programming model (reffered in section ??). The Modelica model exported as FMU for Windows platform is integrated with BOINC wrapper and as a whole it is integrated into BOINC platform deployed on

<sup>15</sup><http://boinc.berkeley.edu/> accessed February 2015

a server as seen in fig. 7.



Obrázek 7: Architecture of parameter sweep application. The whole parameter space is divided into smaller spaces which are resolved by the BOINC workers

The results are described in section ??.

## 11 Results

In previous chapters, there were introduced different methods available for selected use cases in research in biology and medicine. As each of the use cases and available system was proposed on different operating system platform, different architecture and or different middleware the virtualization was utilized to build the virtual infrastructures for purposes of each project. The paper [3] *Infrastructure for Data Storage and Computation in Biomedical Research* in Appendix 15 describes result of establishing the virtualization on physical infrastructure to share computational power among different platforms.

## 12 Conclusion

## 13 Discussion

The result presented in section ?? is an example how a standard format and protocol (DICOM) is utilized to integrate current production system to exchange medical images (MEDIMED) and a grid based solution (Globus MEDICUS) where an underlying technology is hidden for common user. This research was originally motivated by the idea to investigate benefits and show robust grid-based technology against proprietary distributed technology, which may face up to scalability and maintenance issues. Another issue is the philosophy of storing medical images. The presented solution based on the Globus MEDICUS is in general a data warehouse storing one or more copies of DICOM images, in contrast to federated files and metadata stored within home institutions which shares only network infrastructure to interchange the DICOM studies. This seems to be more acceptable by hospitals and by the national and international regulation for clinical and diagnostic use, e.g. The authors of Globus MEDICUS in their further development followed a way of federation of medical images stored within home institutions rather than in a grid infrastructure published by Chervenak et al. [82]. Thus the grid-computing infrastructure for sharing medical images is worth to use in the cases when additional demanding computation e.g. for processing of medical images are needed or for educational, training and knowledge sharing purposes.

The majority of user experience is kept also in the case of the application of remote voice analysis presented in section ?. The processing/analyzing application was kept on it's original platform but moved to remote server and a remote desktop protocol is used to redirect interaction and voice recording from user's computer to remote server and vice versa. Manipulating the . The recordings are stored on remote server for secondary use for further research and analysis algorithm improvement and in case of growth, the long-term storing issue of scientific data will be needed as in the case of medical images. Such kind of service can be deployed on any web server and a occasional need to educate or perform higher number of analysis concurrently can be satisfied with cloud-computing deployment

In the case of application for parameter estimation presented in section ?? the user must fill the data in a form that is a spreadsheet like table and respecting some simple convention. As seen from results, the computation is sensitive on communication overhead therefore more high perfor-

mance computing (HPC) hardware would be beneficial for such computation. However for medium and highly complex models the deployment of worker nodes into cloud-computing environment to form a virtual HPC cluster is worth to consider. The application for parameter sweep is embarrassingly parallel and suitable for high throughput computing (HTC) which is the main focus of current grid-computing infrastructures.

### 13.0.1 Platform

One of the important decision when porting an application to the grid environment is the platform of the used system.

The architecture which involves computational nodes deployed in cloud-computing infrastructure was influenced by the fact, that the model implementation is exported from third party tool to the standard FMU library as mentioned in section 10.3 for the MS Windows platform, which determines the platform of the worker node. The virtualization - or in case of parameter estimation a cloud computing is utilized on prepared platform with MS Windows Datacenter license. In case of parameter sweep a desktop-grid computing BOINC worker and application for MS Windows platform only is prepared for volunteers with the compatible system.

To utilize service-grid infrastructure an export of the model into FMU library and implementation of the wrapper service should be done in the grid-computing platform which is usually Linux based system. Another option could be to use WINE<sup>16</sup> – compatibility layer capable of running Windows applications on several POSIX-compliant operating systems, such as Linux, Mac OSX, BSD.

### 13.0.2 Porting

Each of the introduced systems and application was from its beginning prepared for serial workflow. To achieve higher level of programming model, some manual intervention is usually needed on the system or source code of application.

For the smaller types of application and scientific community with their own tools it is the question, whether to invest on porting their tools to grid specific platform and parallel programming model. In the case of voice science, the analytical application was deployed on virtual machine and made available via remote desktop feature. This caused that users and researchers may stay at their platform and focus on their key technologies and

<sup>16</sup><https://www.winehq.org/> WINE. Accessed March 2015

research rather than to learn new one.

For the case of algorithms, that are already present in grid-computing middleware is key factor the worker/simulation part which is specific for each research community.

## 14 Future work

The "long-tail" movement described by Anderson [83] is business strategy of companies such as Amazon and Apple focusing on offering and delivering not only very popular products but also products with relatively small quantities sold each to final consumers in a price acceptable for them due to reduced sales, marketing and delivery costs. The expansion of Internet and it's related technologies caused this strategy to be profitable and successful and sales of minor products outperforms the most popular products .

There is a discussion about how to preserve the scientific data in a long-term way to prevent loss of them [84, 85] and to facilitate an access to computational resources for large amount of small scientific groups which have limited resources to port, integrate or customize their current tools and processes – the so called long-tail of science. Cloud-computing technologies seems to be customizable and may be an enabling technology to focus on long-tail science consumers as noted e.g. by Weinhardt et al.[86]. The future work in this area may contribute to design and implement policies for long-term scientific data storage and service available for them to gain access to the powerfull capacity of grid-computing and cloud-computing infrastructure.

The medical imaging and processing methods are used to identify the parameters of models of human physiology for further diagnosis statement and treatment decision. E.g. Ralovich et al.[87] proposed a noninvasive method based on computational fluid dynamic and magentic resonance imaging (MRI) to identify pressure difference in aorta for further hemodynamics analysis based on four element windkessel model. However, such type of studies usually focus on particular phenomenon and tries identify parameters of very simplified models that are currently known in systems biology domain. In the time of writing this thesis, Magnetic Particle Imager <sup>17</sup> (available only for animal models and not yet for human medicine) can produce high resolution images with fast shutter speed ( 20 ms). Several computational and storage demanding biomedical application were shown in animal models by Saritas et al.[89]. There are research infrastructures which

---

<sup>17</sup>introduced first by Gleich and Weizenecker[88]



were established to coordinate the research in biology and medicine, e.g. Integrated Structural Biology Infrastructure for Europe (INSTRUCT)<sup>18</sup>, European Life Science Infrastructure for Biological Information (ELIXIR)<sup>19</sup>, European Biomedical Imaging Infrastructure (Euro-BioImaging)<sup>20</sup> and others. As noted by Hunter et al.[90] the purpose of these initiatives is to understand high-level phenotypes from genomic, metabolomic, proteomic, imaging and other types of data which requires multiscale mathematical models and simulation delivered e.g. by virtual physiological human (VPH)<sup>21</sup> project.

The integration with multidimensional models of geometrical, mechanical properties and the time-dependence of the compartments data taken from the medical and biological repositories is a challenge for current "physiome" projects. Current complex models of human physiology are based mainly on lumped parameter approach, some of the compartments are described by physical or chemical laws, while others are only empirically determined curves described by parameters which doesn't have further physiological meaning and can't be measured. Future direction of research may focus on integrative effort of the multiscale modeling which may improve an effort towards patient specific health care, in silico trials and drug discovery.

## 15 Summary

This thesis presents the infrastructure which thanks to virtualization technology joined several domain specific tools in the field of sharing and processing medical images, performing real-time voice analysis and simulating human physiology.

A seamless integration of grid-based PACS system was established with current distributed system to share DICOM medical images. An access to real-time voice analysis application via remote desktop technology brings this type of service to any computer capable to connect to Internet. A system to support analysis and building complex models of human physiology in the phase of parameter estimation and parameter sweep was introduced and additional computational nodes can be flexibly joined by starting prepared virtual machines in cloud-computing deployment. Methodology of building complex models of human physiology was contributed with the idea of acausal and object-oriented modeling techniques.

---

<sup>18</sup><https://www.structuralbiology.eu/> accessed March 2015

<sup>19</sup><http://www.elixir-europe.org/> accessed March 2015

<sup>20</sup><http://www.eurobioimaging.eu/> accessed March 2015

<sup>21</sup><http://www.vph-institute.org/> accessed March 2015

## Reference

- [1] Tomáš Kulhánek and Milan Šárek. Processing of medical images in virtual distributed environment. In *Proceedings of the 2009 Euro American Conference on Telematics and Information Systems: New Opportunities to increase Digital Citizenship*, page 10. ACM, 2009. doi:10.1145/1551722.1551732. 7, 8, 10
- [2] Tomáš Kulhánek, Marek Frič, and Milan Šárek. Remote Analysis of Human Voice–Lossless Sound Recording Redirection. *Analysis of Biomedical Signals and Images. Proceedings of 20th International EURASIP Conference (BIOSIGNAL)*, pages 394–397, 2010. URL: <http://bs2010.biosignal.cz/papers/1092.pdf>. 7, 8, 10
- [3] Tomáš Kulhánek. Infrastructure for data storage and computation in biomedical research. *European Journal for Biomedical Informatics (EJBI)*, 6(1):55–58, 2010. URL: <http://www.ejbi.org/img/ejbi/ejbi2010-1.pdf>. 7, 8, 10, 27, 42
- [4] Tomáš Kulhánek, Marek Mateják, Jan Šilar, and Jiří Kofránek. Parameter estimation of complex mathematical models of human physiology using remote simulation distributed in scientific cloud. In *Biomedical and Health Informatics (BHI), 2014 IEEE-EMBS International Conference on*, pages 712–715, 2014. doi:10.1109/BHI.2014.6864463. 7, 9, 10
- [5] Tomáš Kulhánek, Jiří Kofránek, and Marek Mateják. Modeling of short-term mechanism of arterial pressure control in the cardiovascular system: Object oriented and acausal approach. *Computers in Biology and Medicine*, 54(1):137–144, 2014. doi:10.1016/j.combiomed.2014.08.025. 7, 8, 10, 19
- [6] Tomáš Kulhánek, Martin Tribula, Jiří Kofránek, and Marek Mateják. Simple models of the cardiovascular system for educational and research purposes. *MEFANET Journal*, 2(2):56–63, 2014. 7, 8, 10, 19
- [7] Marek Mateják, Tomáš Kulhánek, and Stanislav Matoušek. Adair-Based Hemoglobin Equilibrium with Oxygen, Carbon Dioxide and Hydrogen Ion Activity. *Scandinavian journal of clinical and laboratory investigation*, 2014. doi:10.3109/00365513.2014.984320. 7, 9, 10
- [8] Tomáš Kulhánek, J Šilar, Marek Mateják, Pavol Privitzer, Jiří Kofranek, and Martin Tribula. Distributed computation and parameter estimation in identification of physiological systems. *VPH conference*, 2010. 7, 9

- [9] Tomáš Kulhánek, Filip Ježek, Marek Mateják, Jan Šilar, Pavol Privitzer, Martin Tribula, and Jiří Kofránek. RESTful web service to build loosely coupled web based simulation of human physiology. *Transactions of Japanese Society for Medical and Biological Engineering*, 51(Supplement):R-32, 2013. doi:[10.11239/jsmbe.51.R-32](https://doi.org/10.11239/jsmbe.51.R-32). 7, 9
- [10] J Kofránek, M Mateják, P Privitzer, M Tribula, T Kulhánek, J Šilar, and R Pecinovský. HumMod-Golem Edition: large scale model of integrative physiology for virtual patient simulators. In *Proceedings of World Congress in Computer Science 2013 (WORLDCOMP'13), International Conference on Modeling, Simulation and Visualisation Methods (MSV'13)*, pages 182–188, 2013. 7, 8, 19
- [11] Marek Mateják, Tomáš Kulhánek, Jan Šilar, Pavol Privitzer, Filip Ježek, and Jiří Kofránek. Physiobrary - Modelica library for Physiology. In *Proceedings of the 10th International Modelica Conference, March 10-12, 2014, Lund, Sweden*, pages 499–505. Linköping University Electronic Press, 2014. doi:[10.3384/ecp14096499](https://doi.org/10.3384/ecp14096499). 7, 8
- [12] Tomáš Kulhánek and Milan Šárek. Virtualizace a integrace v gridovém pacs systému (cze) virtualization and integration in grid pacs system (eng). In *Mefanet 2008*. MSD Brno, 2008. URL: [http://www.mefanet.cz/res/file/mefanet2008/prispevky/25\\_kulhanek.pdf](http://www.mefanet.cz/res/file/mefanet2008/prispevky/25_kulhanek.pdf). 7, 8
- [13] Milan Šárek and Tomáš Kulhánek. Nové směry medicínských aplikací sdružení CESNET. In *sborník příspěvků MEDSOFT*, pages 145–148, 2009. URL: [http://www.creativeconnections.cz/medsoft/2009/Medsoft\\_2009\\_%C5%A0%C3%A1rek\\_Milan.pdf](http://www.creativeconnections.cz/medsoft/2009/Medsoft_2009_%C5%A0%C3%A1rek_Milan.pdf). 7, 8
- [14] Tomáš Kulhánek. Virtual Distributed Environment for Exchange of Medical Images. In D Kuželová, editor, *Proceeding of PhD. Conference 2009 was published as: Doktorandské dny '09*, pages 62–64. Institute of Computer Science/MatfyzPress, 2009. URL: <http://www.cs.cas.cz/hakl/doktorandsky-den/files/2009/sbornik-dd-2009.pdf>. 7, 8
- [15] Tomáš Kulhánek, Marek Frič, and Milan Šárek. Vzdálený přístup k virtuálním výukovým a výzkumným aplikacím - podpora foniatických vyšetření. In Daniel Schwarz, Jaroslav Majerník, Ladislav Dušek, Stanislav Štípek, and Vladimír Mihál, editors, *Mefanet 2009*. MSD Brno, 2009. URL: [http://www.mefanet.cz/res/file/mefanet2009/prispevky/kulhanek\\_full.pdf](http://www.mefanet.cz/res/file/mefanet2009/prispevky/kulhanek_full.pdf). 7

- [16] Tomáš Kulháněk, Marek Frič, and Milan Šárek. Vzdálená analýza lidského hlasu - bezetrátové nahrávání zvuku přes ip síť. In *sborník příspěvků MEDSOFT*, pages 96–101, 2010. 7, 8
- [17] Tomáš Kulháněk, Jan Šilar, Jiří Kofránek, Marek Mateják, Pavol Privitzer, and Martin Tribula. Od výukového modelu k identifikaci fyziologického systému. In Daniel Schwarz, Martin Komenda, Jaroslav Majerník, Stanislav Štípek, Vladimír Mihál, and Ladislav Dušek, editors, *Mefanet 2010*. MSD Brno, 2010. URL: <http://www.mefanet.cz/res/file/mefanet2010/prispevky/kulhanek-full.pdf>. 7
- [18] Tomáš Kulháněk. From Educational Models Towards Identification of Physiological Systems. *Mefanet Report*, 04:69–72, 2011. URL: <http://www.mefanet.cz/res/file/reporty/mefanet-report-2011.pdf>. 7, 9
- [19] Tomáš Kulháněk. Infrastructure for Data Storage and Computation in Biomedical Research. In D Kuželová, editor, *Proceeding of PhD. Conference 2011 was published as: Doktorandské dny '11*, pages 90–94. Institute of Computer Science/MatfyzPress, 2011. URL: <http://www.cs.cas.cz/hakl/doktorandsky-den/files/2011/sbornik-dd-2011.pdf>. 7, 8
- [20] Tomáš Kulháněk, Marek Frič, and Jaroslav Hrb. Vzdálená analýza lidského hlasu v reálném čase. In *sborník příspěvků MEDSOFT*, pages 180–184, 2012. URL: [http://www.creativeconnections.cz/medsoft/2012/Medsoft\\_Kulháněk\\_Tomáš.pdf](http://www.creativeconnections.cz/medsoft/2012/Medsoft_Kulháněk_Tomáš.pdf). 7, 8
- [21] Tomáš Kulháněk, Marek Mateják, Jan Šilar, Pavol Privitzer, Martin Tribula, Filip Ježek, and Jiří Kofránek. HYBRIDNÍ ARCHITEKTURA PRO WEBOVÉ SIMULÁTORY. In *sborník příspěvků MEDSOFT*, pages 115–121, 2013. URL: [http://www.creativeconnections.cz/medsoft/2013/Medsoft\\_2013\\_Kulhanek.pdf](http://www.creativeconnections.cz/medsoft/2013/Medsoft_2013_Kulhanek.pdf). 7, 9
- [22] Tomáš Kulháněk, Marek Mateják, Jan Šilar, and Jiří Kofránek. IDENTIFIKACE FYZIOLOGICKÝCH SYSTÉMŮ. In *sborník příspěvků MEDSOFT*, pages 148–153, 2014. URL: [http://www.creativeconnections.cz/medsoft/2014/Medsoft\\_2014\\_Kulhanek.pdf](http://www.creativeconnections.cz/medsoft/2014/Medsoft_2014_Kulhanek.pdf). 7, 9
- [23] Tomáš Kulháněk and Marek Frič. Remote analysis of human voice - environment for voice training and ORL medicine. In *EGI Technical Forum, Prague*, Prague, 2012. 7
- [24] Jiří Kofránek, Pavol Privitzer, Martin Tribula, Jan Šilar, Tomáš Kulháněk, Marek Mateják, and Filip Ježek. Virtual patient simulator, CZ

- Utility model. Application number: 2014-30329. Issue number: 27613, 2014. URL: <http://isdv.upv.cz/portal/pls/portal/portlets.pts.det?xprim=10082946>. 7, 9
- [25] Issac N. Bankman, editor. *Handbook of Medical Imaging; Processing and Analysis*. 2000. 12
- [26] K. Slavicek, O. Dostal, M. Javornik, and M. Drdla. MEDIMED - Regional Centre for Medicine Image Data Processing. *2010 Third International Conference on Knowledge Discovery and Data Mining*, pages 310–313, January 2010. doi:[10.1109/WKDD.2010.133](https://doi.org/10.1109/WKDD.2010.133). 14, 15
- [27] Peeter Ross, Ruth Sepper, and Hanna Pohjonen. Cross-border teleradiology-experience from two international teleradiology projects. *European journal of radiology*, 73(1):20–5, January 2010. doi:[10.1016/j.ejrad.2009.10.016](https://doi.org/10.1016/j.ejrad.2009.10.016). 14
- [28] Vanessa Saliba, Helena Legido-Quigley, Riina Hallik, Ain Aaviksoo, Josip Car, and Martin McKee. Telemedicine across borders: a systematic review of factors that hinder or support implementation. *International journal of medical informatics*, 81(12):793–809, December 2012. doi:[10.1016/j.ijmedinf.2012.08.003](https://doi.org/10.1016/j.ijmedinf.2012.08.003). 14
- [29] S. G. Erberich, M. Bhandekar, M. D. Nelson, a. Chervenak, and C. Kesselman. DICOM grid interface service for clinical and research PACS: A globus toolkit web service for medical data grids. *International Journal of Computer Assisted Radiology and Surgery*, 1(ii):100–102, 2006. doi:[10.1007/s11548-006-0013-0](https://doi.org/10.1007/s11548-006-0013-0). 14, 15
- [30] Stephan G Erberich, Jonathan C Silverstein, Ann Chervenak, Robert Schuler, Marvin D Nelson, and Carl Kesselman. Globus MEDICUS - federation of DICOM medical imaging devices into healthcare Grids. *Studies in health technology and informatics*, 126:269–278, 2007. 14, 15
- [31] H Duque, J Montagnat, J Pierson, L Brunie, and I E Magnin. DM 2 : A Distributed Medical Data Manager for Grids. In *Biogrid'03 workshop*, pages 138–147, Tokyo, Japan. 14
- [32] Johan Montagnat, Ákos Frohner, Daniel Jouvenot, Christophe Pera, Peter Kunszt, et al. A Secure Grid Medical Data Manager Interfaced to the gLite Middleware. *Journal of Grid Computing*, 6(1):45–59, July 2007. doi:[10.1007/s10723-007-9088-2](https://doi.org/10.1007/s10723-007-9088-2). 14

- [33] Dagmar Krefting, Julian Bart, Kamen Beronov, Olga Dzhimova, Jürgen Falkner, et al. MediGRID: Towards a user friendly secured grid infrastructure. *Future Generation Computer Systems*, 25(3):326–336, 2009. doi:[10.1016/j.future.2008.05.005](https://doi.org/10.1016/j.future.2008.05.005). 14
- [34] Dagmar Krefting, Michal Vossberg, Andreas Hoheisel, and Thomas Toxdorff. Simplified implementation of medical image processing algorithms into a grid using a workflow management system. *Future Generation Computer Systems*, 26(4):681–684, 2010. doi:[10.1016/j.future.2009.07.004](https://doi.org/10.1016/j.future.2009.07.004). 14
- [35] K Skaburskas, F Estrella, J Shade, D Manset, J Revillard, et al. Health-e-Child: a grid platform for european paediatrics. *Journal of Physics: Conference Series*, 119:082011, 2008. doi:[10.1088/1742-6596/119/8/082011](https://doi.org/10.1088/1742-6596/119/8/082011). 14
- [36] Siegfried Benkner, Antonio Arbona, Guntram Berti, Alessandro Chiarini, Robert Dunlop, et al. @neurIST: Infrastructure for advanced disease management through integration of heterogeneous data, computing, and complex processing services. *IEEE Transactions on Information Technology in Biomedicine*, 14(6):1365–1377, 2010. doi:[10.1109/TITB.2010.2049268](https://doi.org/10.1109/TITB.2010.2049268). 14
- [37] Martin Kuba, Ondrej Krajíček, Petr Lesný, Jan Vejvalka, and Tomáš Holeček. Grid empowered sharing of medical expertise. *Studies in health technology and informatics*, 120:273–282, 2006. 14
- [38] Martin Kuba, Daniel Kouřil, and Michal Procházka. Sealed grid with downloadable services. In Thierry Priol and Marco Vanneschi, editors, *Towards Next Generation Grids*, pages 83–91. Springer US, 2007. doi:[10.1007/978-0-387-72498-0\\_8](https://doi.org/10.1007/978-0-387-72498-0_8). 14
- [39] I Foster. Globus toolkit version 4:Software for service-oriented systems. *Journal of Computer Science and Technology*, 21(4):513–520, 2006. 15
- [40] William Allcock, John Bresnahan, Rajkumar Kettimuthu, Michael Link, Catalin Dumitrescu, Ioan Raicu, and Ian Foster. The Globus striped GridFTP framework and server. *Proceedings of the ACM/IEEE 2005 Supercomputing Conference, SC’05*, 2005(c):1–11, 2005. doi:[10.1109/SC.2005.72](https://doi.org/10.1109/SC.2005.72). 15
- [41] JW Cooley and JW Tukey. An algorithm for the machine calculation of complex Fourier series. *Mathematics of computation*, pages 297–301, 1965. URL: <http://www.jstor.org/stable/2003354>. 16

- [42] M. Frigo and S.G. Johnson. The Design and Implementation of FFTW3. *Proceedings of the IEEE*, 93(2):216–231, 2005. URL: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1386650>, doi:10.1109/JPROC.2004.840301. 16
- [43] Wendy DeLeo LeBorgne and Barbara D. Weinrich. Phonetogram changes for trained singers over a nine-month period of vocal training. *Journal of Voice*, 16(1):37–43, 2002. doi:10.1016/S0892-1997(02)00070-X. 16
- [44] FLORIS L Wuyts, LOUIS Heylen, FONS Mertens, M D Caju, RAOUL Rooman, P H de Heyning, and Marc De Bodt. Effects of age, sex, and disorder on voice range profile characteristics of 230 children. *ANNALS OF OTOLOGY RHINOLOGY AND LARYNGOLOGY*, 112(6):540–548, 2003. 16
- [45] Jan G. Švec and Harm K. Schutte. Videokymography: High-speed line scanning of vocal fold vibration. *Journal of Voice*, 10(2):201–205, 1996. doi:10.1016/S0892-1997(96)80047-6. 16
- [46] Jan G Svec, Frantisek Sram, and Harm K Schutte. Videokymography in voice disorders: what to look for? *The Annals of otology, rhinology, and laryngology*, 116(3):172–80, March 2007. URL: <http://europepmc.org/abstract/MED/17419520>. 16
- [47] Marek Frič. Parametrizovaný fonetogram obecných řečových a hlasových projevů - ParVRP, 2007. URL: <http://zvuk.hamu.cz/vyzkum/dokumenty/TL12x.pdf>. 16, 17
- [48] Marek Frič, Tomáš Kulhánek, and Jaroslav Hrb. Systém pro vzdáleně přístupnou analýzu hlasu RealVoiceLab, 2012. URL: <http://zvuk.hamu.cz/vyzkum/dokumenty/TL46x.pdf>. 16, 17
- [49] P Kohl, E J Crampin, T A Quinn, and D Noble. Systems biology: an approach. *Clinical pharmacology and therapeutics*, 88:25–33, 2010. doi:10.1038/clpt.2010.92. 18
- [50] P. J. Hunter and M. Viceconti. The VPH-Physiome Project: Standards and Tools for Multiscale Modeling in Clinical Applications. *IEEE Reviews in Biomedical Engineering*, 2:40–53, 2009. doi:10.1109/RBME.2009.2036204. 18

- [51] A C Guyton, T G Coleman, and H J Granger. Circulation: overall regulation. *Annual review of physiology*, 34:13–46, 1972. doi:10.1146/annurev.ph.34.030172.000305. 18
- [52] Robert L Hester, Radu Iliescu, Richard Summers, and Thomas G Coleman. Systems biology and integrative physiological modelling. *The Journal of physiology*, 589(Pt 5):1053–60, March 2011. doi:10.1113/jphysiol.2010.201558. 18
- [53] Robert L Hester, Alison J Brown, Leland Husband, Radu Iliescu, Drew Pruett, Richard Summers, and Thomas G Coleman. HumMod: a modeling environment for the simulation of integrative human physiology. *Frontiers in physiology*, 2(April):12, January 2011. doi:10.3389/fphys.2011.00012. 18
- [54] Erik Butterworth, Bartholomew E. Jardine, Gary M. Raymond, Maxwell L. Neal, and James B. Basingthwaight. JSim, an open-source modeling system for data analysis. *F1000Research*, 2, May 2014. doi:10.12688/f1000research.2-288.v2. 18
- [55] P J Hunter. The IUPS Physiome Project: a framework for computational physiology. *Progress in biophysics and molecular biology*, 85(2-3):551–69, January 2004. doi:10.1016/j.pbiomolbio.2004.02.006. 18
- [56] Tommy Yu, Catherine M Lloyd, David P Nickerson, Michael T Coling, Andrew K Miller, et al. The Physiome Model Repository 2. *Bioinformatics (Oxford, England)*, 27(5):743–744, 2011. doi:10.1093/bioinformatics/btq723. 18
- [57] Michael Hucka, ABBJ Finney, Benjamin J Bornstein, Sarah M Keating, Bruce E Shapiro, et al. Evolving a lingua franca and associated software infrastructure for computational systems biology: the Systems Biology Markup Language (SBML) project. *Systems biology*, 1(1):41–53, 2004. 19
- [58] Nicolas Le Novère, Benjamin Bornstein, Alexander Broicher, Mélanie Courtot, Marco Donizelli, et al. BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic acids research*, 34:D689–D691, 2006. doi:10.1093/nar/gkj092. 19
- [59] J Kofránek and J Rusz. Restoration of Guyton’s diagram for regulation of the circulation as a basis for quantitative physiological model development. *Physiological research / Academia Scientiarum Bohemoslovaca*, 59(6):897–908, January 2010. 19



- [60] J Kofranek, M Matejak, and P Privitzer. Hummod-large scale physiological models in modelica. In *Proceedings 8th Modelica Conference, Dresden, Germany*, pages 713–724, 2011. 19
- [61] Javier Fernandez de Canete, P del Saz-Orozco, D Moreno-Boza, and E Duran-Venegas. Object-oriented modeling and simulation of the closed loop cardiovascular system by using SIMSCAPE. *Computers in biology and medicine*, 43(4):323–33, May 2013. doi:10.1016/j.compbimed.2013.01.007. 19
- [62] J Fernandez de Canete, J Luque, J Barbancho, and V Munoz. Modeling of long-term and short-term mechanisms of arterial pressure control in the cardiovascular system: An object-oriented approach. *Computers in biology and medicine*, 47:104–112, April 2014. doi:10.1016/j.compbimed.2014.01.006. 19
- [63] Peter Fritzson and Peter Bunus. Modelica-a general object-oriented language for continuous and discrete-event system modeling and simulation. In *Simulation Symposium, 2002. Proceedings. 35th Annual*, pages 365–380. IEEE, 2002. 19
- [64] Michael Tiller. Modelica by Example, 2014. URL: <http://book.xogeny.com>. 19
- [65] M. Hofmann. On the Complexity of Parameter Calibration in Simulation Models. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, 2(4):217–226, October 2005. doi:10.1177/154851290500200405. 20
- [66] P Eykhoff. *Trends and progress in system identification*. IFAC series for graduates, research workers and practising engineers;vol. 1. Pergamon, Oxford, 1981. 20
- [67] Michael C K Khoo. *Physiological control systems*. IEEE press, 2000. 20
- [68] Carmen G Moles, Pedro Mendes, and Julio R Banga. Parameter Estimation in Biochemical Pathways: A Comparison of Global Optimization Methods. *Genome Research*, pages 2467–2474, 2003. doi:10.1101/gr.1262503. 20
- [69] William A Pruett, Leland D Husband, Graham Husband, Muhammad Dakhalla, Kyle Bellamy, Thomas G Coleman, and Robert L Hester. A Population Model of Integrative Cardiovascular Physiology. *PloS one*, 8(9):e74329, 2013. 20

- [70] Shuji Takahashi, Kenta Sakawa, Yoichi Shiraishi, Hiroshi Miyashita, and Science Technology. Modeling, simulation and parameter estimation of the cardiovascular system by using Model Based approach. pages 493–500, 2013. [20](#)
- [71] Mohamed a. Abbass, Emad ElSamahy, and Ahmed Genedy. An optimized cardiovascular model: A pattern search approach. *Proceedings - 2012 International Conference on Biomedical Engineering and Biotechnology, iCBEB 2012*, 1:45–48, 2012. [doi:10.1109/iCBEB.2012.62](#). [20](#)
- [72] Sergio Maffioletti, Riccardo Murri, B Jonen, and S Scheuring. Computational workflows with GC3Pie. In *Proceedings of the EGI Community Forum*, 2012. [20](#)
- [73] Marty Humphrey, Norm Beekwilder, Jonathan L. Goodall, and Mehmet B. Ercan. Calibration of watershed models using cloud computing. *2012 IEEE 8th International Conference on E-Science*, (1048222):1–8, October 2012. [doi:10.1109/eScience.2012.6404420](#). [20](#)
- [74] Bryan a. Tolson and Christine a. Shoemaker. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research*, 43(August 2006):1–16, 2007. [doi:10.1029/2005WR004723](#). [20](#)
- [75] John H Holland. *Adaptation in Natural and Artificial Systems*, volume Ann Arbor. 1975. URL: <http://www.citeulike.org/group/664/article/400721>. [21](#)
- [76] T Blochwitz, M Otter, M Arnold, and C Bausch. The functional mockup interface for tool independent exchange of simulation models. In *Proceeding of 8th Modelica Conference, Dresden*, 2011. [21](#)
- [77] J. C. Helton, J. D. Johnson, C. J. Sallaberry, and C. B. Storlie. Survey of sampling-based methods for uncertainty and sensitivity analysis. *Reliability Engineering and System Safety*, 91:1175–1209, 2006. [doi:10.1016/j.res.2005.11.017](#). [23](#)
- [78] Andrea Saltelli, Stefano Tarantola, Francesca Campolongo, and Marco Ratto. *Sensitivity Analysis in Practice*. 2004. [doi:10.1002/0470870958](#). [23](#)
- [79] Andrea Saltelli, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana, and Stefano Tarantola. *Global Sensitivity Analysis. The Primer*. 2008. URL: [http://](#)

- [eu.wiley.com/WileyCDA/WileyTitle/productCd-0470059974.html](http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470059974.html), doi:10.1002/9780470725184. 23
- [80] Peter Kacsuk. P-GRADE portal family for grid infrastructures. *Concurrency Computation Practice and Experience*, 23:235–245, 2011. doi:10.1002/cpe.1654. 23
- [81] DP Anderson. Boinc: A system for public-resource computing and storage. In *Grid Computing, 2004. Proceedings. Fifth IEEE/ACM International Workshop on*. IEEE, 2004. doi:10.1109/GRID.2004.14. 23
- [82] Ann L. Chervenak, Theo G M Van Erp, Carl Kesselman, Mike D’Arcy, Janet Sobell, et al. A system architecture for sharing de-identified, research-ready brain scans and health information across clinical imaging centers. *Studies in Health Technology and Informatics*, 175:19–28, 2012. doi:10.3233/978-1-61499-054-3-19. 28
- [83] Chris Anderson. The Long Tail: How Endless Choice Is Creating Unlimited Demand. *Word Journal Of The International Linguistic Association*, page 256, 2006. URL: <http://www.amazon.co.uk/dp/184413850X>, doi:30. 29
- [84] Timothy H. Vines, Arianne Y K Albert, Rose L. Andrew, Florence Débarre, Dan G. Bock, et al. The availability of research data declines rapidly with article age. *Current Biology*, 24:94–97, 2014. arXiv:1312.5670, doi:10.1016/j.cub.2013.11.014. 30
- [85] P. Bryan Heidorn. Shedding Light on the Dark Data in the Long Tail of Science. *Library Trends*, 57(2):280–299, 2008. doi:10.1353/lib.0.0036. 30
- [86] Christof Weinhardt, Arun Anandasivam, Benjamin Blau, Nikolay Borissov, Thomas Meinl, Wibke Michalk, and Jochen Stöß er. Cloud Computing – A Classification, Business Models, and Research Directions. *Business & Information Systems Engineering*, 1:391–399, 2009. doi:10.1007/s12599-009-0071-2. 30
- [87] Kristóf Ralovich, Lucian Itu, Viorel Mihalef, Puneet Sharma, Razvan Ionasec, et al. Hemodynamic assessment of pre- and post-operative aortic coarctation from MRI. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 15:486–93, 2012. URL: <http://www.ncbi.nlm.nih.gov/pubmed/23286084>. 30

- [88] Bernhard Gleich and Jürgen Weizenecker. Tomographic imaging using the nonlinear response of magnetic particles. *Nature*, 435(7046):1214–7, June 2005. URL: <http://www.nature.com.ezproxy.techlib.cz/nature/journal/v435/n7046/full/nature03808.html>, doi:10.1038/nature03808. 30
- [89] Emine U. Saritas, Patrick W. Goodwill, Laura R. Croft, Justin J. Konkle, Kuan Lu, Bo Zheng, and Steven M. Conolly. Magnetic particle imaging (MPI) for NMR and MRI researchers. *Journal of Magnetic Resonance*, 229:116–126, 2013. URL: <http://dx.doi.org/10.1016/j.jmr.2012.11.029>, doi:10.1016/j.jmr.2012.11.029. 30
- [90] Peter Hunter, Tara Chapman, Peter V Coveney, Bernard de Bono, Vanessa Diaz, et al. A vision and strategy for the virtual physiological human: 2012 update. *Interface Focus*, 3(2), February 2013. URL: <http://rsfs.royalsocietypublishing.org/content/3/2/20130004.abstract>. 30

Infrastructure for data storage and computation in biomedical research  
The paper [3] published as

Tomáš Kulhánek. Infrastructure for data storage and computation in biomedical research. *European Journal for Biomedical Informatics (EJBI)*, 6(1):55–58, 2010. URL: <http://www.ejbi.org/img/ejbi/ejbi2010-1.pdf>