

# Chapter 4

## Systems Biology, Information Technology, and Cancer Research

Imme Petersen, Regine Kollek, Anne Brüninghaus, and Martin Döring

**Abstract** The plethora and heterogeneity of data on biological processes have caused a change in approaches to data handling and processing by using high-performance computing and informatics. Infrastructures based on information and communication technology (ICT) have been developed to facilitate data management, access, and sharing of data on biological structures and processes on which systems biology is based. Although such infrastructures are essential for research and collaboration, they are often not regarded as being part of knowledge production. In contrast to this, we hypothesize that ICT infrastructures are not mere service facilities to support research activities, but enable, and restrict doing systems research at the same time. Based on a case study in systems cancer research, we argue that the understanding and modeling of biological systems is profoundly shaped by ICT and their underlying conceptualizations. In addition, individual scientists and research institutions cede the responsibilities of the activities associated with standardization, integration, and management of data. From the perspective of the sociological Actor-Network-Theory, our analysis also showed that such ICT infrastructures will become new powerful actors for knowledge production and within the knowledge-producing community of systems biology. Individual scientists and research institutions often neglect the challenges related to standardization, integration, and management of data that complicates and sometimes impedes innovation and translation of new developments into practice. This implies that standardization and integration in systems biology are as important as data generation.

**Keywords** Scientific practice • ICT infrastructure • Data management • Integration • Standardization • Case study

The development of systems biology has only been possible through the application of information and communication technology (ICT) to handle the large volume and variety of data about molecular processes in cells and organisms. Databases and

---

I. Petersen (✉) • R. Kollek • A. Brüninghaus • M. Döring  
Research Centre for Biotechnology, Society and the Environment (FSP BIOGUM),  
University of Hamburg, Hamburg, Germany  
e-mail: [imme.petersen@uni-hamburg.de](mailto:imme.petersen@uni-hamburg.de)

infrastructures based on ICT were established for systems biological research to support systematization and integration of data on genomes, transcriptomes, and proteomes. To understand better the practices of data handling and data use in systems biology or, in a broader sense, to get an idea of systems biology in the making, we investigate how systems-oriented research is organized and performed in an ICT-based research environment. Even though ICT infrastructures are often considered as service facilities to ease research, we hypothesize that the understanding and modeling of biological systems is deeply shaped by ICT and their underlying design and conceptualization. Our second hypothesis is that the application of ICT enables and restricts doing systems research at the same time.

Therefore, we were looking for influences, dependencies, translations, and potentials that have occurred when systems biology has met ICT for doing research. As far as we are aware, little is known about the complexity and dynamics of the relationship between systems research and information technology. We therefore decided to use an exploratory research strategy known as case study. By carefully describing an individual case, the study aims at giving a deep insight into the subject of the chosen case and at drawing indications from it for further hypothesis creation on the subject. We empirically explored the challenges of organizing and doing systems-oriented research in an ICT environment in the applied field of systems medicine. In general, systems medicine is the making of systems biology in the making, as it implements systems biological approaches in medical concepts, research, and practice. The case under study was an international research project in which an integrated European ICT infrastructure was designed and developed in support of the systems-oriented research community in oncology.

After a short introduction into the case under study and the methods used in the empirical analysis, we present the empirical results in regard to the needs and demands of coordinating data collections in a computational environment (Sect. 4.1). In the second part, we trace the development of *in silico* oncology to understand better the underlying ideas regarding data analysis in systems-oriented research in cancer (Sect. 4.2). In focus is a knowledge discovery tool called the oncosimulator that was built in the course of the project. Based on the analysis of the results and outcomes of the case study, we retrace the current status of ICT in systems-oriented research and assess the potential of such an approach (Sect. 4.3). In the last section, we discuss what kind of function and role ICT infrastructures may in fact play in systems-oriented research in oncology in the future (Sect. 4.4).

## **4.1 Computers, Cancers, and Clinics: Coordinating Systems-Oriented Research in Oncology**

High-throughput production of genomic data has confirmed that cancer can be regarded as a system: moving from single gene-based molecular investigation to molecular network research is seen as the most promising track to discover the mechanistic underpinning of cancer. It is assumed that cancer generally arises from

disease-perturbed networks and that different network perturbations lead to different cancers (Lin et al. 2005). Additionally, it seems likely that network perturbations change with cancer progression. Such cancer-related, perturbed networks are currently under study to understand better how the cancer genome functions as a complex biological system in individual patients. The shift of interest from the identification of individual cancer components to the ways that these components interact has led to an explosion in the number of different types of data generated from the patients. The following data types are acquired.

- Molecular data types often referred to as Omics data (e.g., DNA variations, RNA, proteins, metabolites)
- Epigenetic data (e.g., DNA methylation patterns)
- Clinical data collected on clinical case report forms (e.g., symptoms, histology, administered treatment, treatment response)
- Imaging data (e.g., MRI, CT, ultrasound)
- Pathology data and other laboratory data

In order to understand better or even intervene more effectively in cancer and its development, these different data types are assembled. How these components involved in the processes under investigation might relate to and react with each other is systematically explored and formalized in mathematical models (e.g., Wolkenhauer and Green 2013). In this sense, data integration describes a dynamic process in which different data types and methods as well as disciplinary explanations and approaches are combined (O'Malley and Soyer 2012, 59).

The data types that are collected to build models of formalized relations and interactions are usually managed and stored in separate databanks in different geographical sites. Integration of data coming from such databanks raises, however, questions concerning data protection and audited data access.<sup>1</sup> Further problems have occurred: first, data coming from different sources vary significantly in terms of the contexts and circumstances they were gathered and stored (e.g., location, national law, history of data collection, context of application); second, efficient integration of different data sets is often also hampered by conflicting terminology and classification (e.g., Meier and Gehring 2008).

In regard to Omics data, many external technological platforms aim at solving this problem. They offer quality assessment of single Omics data types, which is needed to control the variations in the large number of biological and experimental parameters involved in data production. For example, to profile DNA methylation at least five different techniques exist that capture slightly different aspects of this process referring to the epigenetic modification or reprogramming of DNA (Rakyan et al. 2011). Accordingly, it is still a bioinformatic challenge to analyze a single type of Omics data because there are different approaches to raise data and different platforms to safeguard them.

The hurdles or even inability to share data and technologies is considered to be a bottleneck of the research process as it hampers efficient research collaborations

---

<sup>1</sup>For further information on data protection see, for example, Forgó et al. (2010).

(Swertz and Jansen 2007). Researchers even have problems integrating data from different technologies within a single laboratory. As a result, “clinicians or molecular biologists often find it hard to exploit each other’s expertise due to the absence of a cooperative environment which enables the sharing of data, resources or tools for comparing results and experiments, and a uniform platform supporting the seamless integration and analysis of disease-related data at all levels” (Tsiknakis et al. 2006, 248).

One of the major challenges of systems medical research is, however, to translate laboratory findings into clinical treatments. It aims at finding ways to tailor therapy to the molecular characteristics of individual patients that can be used for precise diagnosis and as targets for novel treatments. Hence, the patient’s biological profile arising from different molecular techniques has to be combined with clinical data relevant to the development, treatment, and prognosis of cancer (Abu-Asab et al. 2013). As of today, there is still no common methodology for integrating data types such as genomic and clinical data or proteomic and imaging data (Green and Wolkenhauer 2012). However, in translational research a few success stories already exist. For instance, gene expression profiling was used to classify tumors into subgroups representing distinct disease states that respond differently to currently used therapies. Finally, these experiments were successful in predicting the likelihood of chemotherapy benefit for patients with low-grade breast cancer and to quantify the likelihood of recurrence (Symmans et al. 2010; Lee et al. 2010; Desmedt et al. 2011).

### ***4.1.1 Description of the Case Study and the Empirical Approach***

After introducing systems-oriented research in cancer as the field of study, we now give a short description of the case under study and the empirical approach to study the scientific practice of systems biology.

#### **4.1.1.1 Case Study**

To elaborate on the relationship between systems-oriented research and information technology we chose an exploratory research strategy known as case study. The case—which can be persons, events, decisions, periods, projects, or institutions—is empirically inquired within its real-life context by using multiple sources and one or more methods (Thomas 2011). By intensely looking at an individual case, conclusions can only be drawn about that case in its specific context. From this it follows that emphasis is placed on exploration and description and not on testing generalizable hypotheses. However, case studies aim at giving deep insights into the subject of the chosen case and drawing indications from it to allow further elaboration and hypothesis creation on the subject (Yin 2009).

This is why we chose to analyze empirically the conception and realization of an ICT infrastructure in the domain of cancer research (see Box 4.1). The case under study is the research project “ACGT—Advancing Clinico-Genomic Trials on Cancer: Open Grid Services for Improving Medical Knowledge Discovery” funded by the 6th Framework Program of the European Commission under the Action Line “Integrated Biomedical Information for Better Health” (FP6/2004/IST-026996). From February 2006 until July 2010, 26 research groups from 12 European countries and Japan designed and developed an integrated technological platform in support of postgenomic, multicentric clinical trials targeting two major cancer diseases, namely, breast cancer and pediatric nephroblastoma, a childhood cancer of the kidneys (see Box 4.2).

**Box 4.1: Distinction Between and Meanings of ICT Infrastructure, Platform, Architecture, and Environment**

In this chapter, we use the terms *infrastructure*, *platform*, *architecture*, and *environment* to emphasize different meanings of information and communication technology (ICT) regarding the case study. In order to be as clear as possible in our terminology, we use the following definitions.

**ICT infrastructure:** The term describes the technology as a new phenomenon in science. The ACGT infrastructure is an example for the phenomenon having individual characteristics.

**ICT platform:** The term emphasizes the utilization of an ICT infrastructure by the users.

**ICT architecture:** The term refers to the technological components of a distinct infrastructure and how these components are interconnected in technological terms.

**ICT environment:** The term emphasizes the broader context of how ICT is integrated into science not focusing on an individual case.

**Box 4.2: Definitions of Genomics, Postgenomics, Molecular Technologies, and Clinical Trial**

Again, in order to be as clear as possible in our terminology, we use the following definitions.

**Genomics/postgenomics:** *Genomics* is part of genetics that applies recombinant DNA, DNA sequencing methods, and bioinformatics to sequence, assemble, and analyze the structure and the function of genes and genomes and studies their expression and regulation. *Postgenomics* refers to any fields of study that is only possible after the genome of an organism is published.

(continued)

Postgenomic research investigates which genes are active at particular times and under different environmental conditions (gene expression), for example, how genes are transcribed into messenger RNA, the chemical that carries the instructions for forming proteins (transcriptomics), how genes are expressed as proteins (proteomics), and how they influence the chemicals that control our cellular biochemistry and metabolism (metabolomics).

**Molecular technologies:** Molecular technologies are used to characterize, isolate, and manipulate the molecular components of cells and organisms. Thus, molecular technologies are the basic tools to study genetic information. Polymerase chain reaction (PCR) is the most basic molecular technology. It is used to produce multiple identical copies of DNA fragments. Other key technologies include DNA sequencing methods used to determine the order of the four bases (Adenine, Guanine, Cytosine, Thymine) in a strand of DNA and DNA microarrays that visualize the gene expression of an organism at a particular stage (expression profiling).

**Clinical trial/study:** Clinical trials are prospective biomedical or behavioral research studies on patients or volunteers that are designed to answer specific questions about biomedical or behavioral interventions, such as drugs, vaccines, biological products, surgery procedures, radiology procedures, devices, behavioral treatments, process-of-care changes, or preventive care. *Multicentric trials* are conducted in several locations (e.g., clinical centers). *Clinico-genomic trials* explicitly approach the integration of genomic data with clinical data in medical research.

The ultimate initial aims of the ACGT consortium were: (1) to design experiments for obtaining coherent and consistent medical and biological data, while avoiding various types of biases and errors; (2) to develop methods for integrating heterogeneous (e.g., genomic, medical) data sources, including the use of ontologies that facilitate mapping and information retrieval; (3) to develop methods for selection, checking, cleaning, and pre-processing of combined genomic-medical data; and (4) to incorporate collaborative approaches to data analysis, inasmuch as biomedical statisticians and data miners in genomics and medicine have been following different methodologies and dedicated, often proprietary, tools (ACGT 2005, 9).<sup>2</sup>

To address these different goals, different tools and services were developed and implemented in the ACGT platform. The main technical components of the ACGT infrastructure are the following (ACGT 2005, 11f).

---

<sup>2</sup> As two of the authors were part of the ACGT consortium, we had access to unpublished working papers, for example, ACGT (2005) Annex 1—Description of Work, Proposal.

- *Biomedical technology Grid layer.* The Grid technology comprises the basic technology for scheduling and brokering of resources. This layer-based architecture offers seamless mediation services for sharing data and data processing methods and tools, and advanced security tools according to European legal and ethical regulations.
- *Distributed data access and applications.* A set of compatible software services based on Web services provide uniform data access to distributed and heterogeneous data sources, that is, clinical data, eHealth records, microarrays, SNP data, and the like.
- *Ontologies and semantic mediation tools.* Formalized knowledge representations (ontologies) are required to facilitate semantic data integration as well as annotation and data analysis of large-scale biomedical data. The ACGT infrastructure offers a reference ontology for the field targeted by the ACGT project.
- *Clinical trial management system.* The clinical trial builder based on an ontology-driven software aims at helping to set up new clinico-genomic trials easily, to collect different types of data, and to put researchers in the position to perform cross-trial analysis.
- *Technologies and tools for in silico oncology.* The oncosimulator models tumor growth and therapy response in silico. The aim here is to create patient-specific computer simulation models of the biological activity of malignant tumors and normal tissues in order to optimize the therapeutic schemes and to contribute to the understanding of the disease at the molecular, cellular, and higher levels of complexity.
- *Grid-enabled application layer.* The data-mining Grid services support and improve complex knowledge discovery processes and knowledge extraction operations. Ways are sought for enabling easy integration and reuse of existing bioinformatics services into the ACGT infrastructure. Analytic services, for example, literature mining, visualization of results, and so on are also been implemented.
- *The integrated ACGT architecture.* Integration of applications requires a composite service that orchestrates other services in order to interoperate in a workflow. The ACGT workflow editor organizes and ensures that data formats are compatible and that semantic relationships between objects shared or transferred in workflows are clear. This creates any easy-to-use workflow environment so that researchers can design their discovery workflows.

The project was structured into a number of interrelated milestones representing the different components and tasks: (1) user requirement analysis and specifications; (2) technologies and services; (3) trust and security; (4) clinical trial implementation, verification, and demonstration; and (5) project management (ACGT 2005, 49f). The major milestones of the project representing the achievement of the objectives and goals were assigned to one of the 16 work packages. Work package (WP) 1 was responsible for the formation of the management structure of the project and the coordination of the project activities. WP 2 explored the user needs and requirements. WP 3 and WP 4 were assigned to build the Grid architecture; in WP 5 the data

access services were constructed, and in WP 6 data mining and knowledge discovery tools were developed. The ontologies for mediation services and the clinical trials and applications were assigned to WP 7 and technologies and tools for *in silico* oncology were built in WP 8. The ethical and legal requirements have been addressed to WP 10 and the security requirements to WP 11, and a separate work package (WP 9) integrated the various tools to synthesize the integrated ACGT architecture. The evaluation and validation of the infrastructure was done in WP 13, the training aspects arising from clinico-genomic integration in WP 14 and the dissemination in WP 15. Last, WP 16 was responsible for the market investigation and the exploitation plan of ACGT (ACGT 2005, 121).

At the end of the ACGT research project, a first prototype of an ICT infrastructure was delivered that facilitated the integrated and secure access to heterogeneous data sources (e.g., distributed clinical trial databases). Furthermore, the ACGT prototype provided a range of reusable, open source analytical tools for the analysis of such integrated, multilevel clinico-genomic data. The data analyses were supported by discovery-driven analytical workflows. Finally, these research activities complied with existing ethical and legal regulation (ACGT 2005, 10; Bucur et al. 2011, 1120).

#### 4.1.1.2 Empirical Approach

Our empirical approach was stimulated by and based upon the fact that two of the authors (Regine Kollek and Imme Petersen) participated in the ACGT project as consortium members from February 2006 until July 2010. Coordinating the ethical framework, we collaborated in particular with the partners being responsible for legal and security requirements in WP 10 and 11 as well as with the clinical partners in WP 12 and attended the consortium meetings taking place every 6 months. After the project ended in July 2010, we conducted guided interviews with selected project participants.

To select interview partners, we wanted first to identify the most relevant actors within the ACGT consortium keeping the consortium and the project running. We assumed that the ACGT consortium was a network of actors working together on the joint task of developing an integrative ICT infrastructure. Accordingly, the most relevant actors were the ones working most intensely in cooperation with other ACGT participants. As publications in such large research projects are usually based on joint work, we conducted a bibliometric analysis of the collaboration for internal publications (deliverables) and external publication (peer-reviewed articles, books, conference proceedings).<sup>3</sup>

As a first step, we identified the ACGT participants and counted their coauthorships for deliverables. Deliverables are usually created within the work packages; however, we wanted to cover all kinds of cooperation. Therefore, we counted the amount of cooperation per coauthor across work packages as well. We summed up

---

<sup>3</sup>Many of the deliverables are still available online. ACGT, <http://acgt.ercim.eu/documents/public-deliverables.html>. Accessed December 11, 2014.

the amount of coauthorships for individual actors and added to it the amount of cooperation across work packages. Resulting from that, we had a data cluster that comprised coauthorship and cooperation within the ACGT consortium. Additionally, we checked if all actors having designated tasks within the project (e.g., work package leadership, project management, quality control) were included in our sample. This was the case. The first most active 20 project participants were chosen; 18 scientists consented to an interview (13 computer scientists (IT), 4 biomedical researchers such as biologists, biostatisticians, and clinicians (BioMed), and 1 lawyer (LAW). They were queried using a theme-structured interview guideline. The interviews were focused on the participants' personal experiences as well as their judgments regarding the ACGT project. The interview guideline was structured into four sections addressing the following topics: (1) experiences of scientific and practical cooperation in the ACGT project (in particular, interdisciplinary negotiations); (2) experiences regarding the realization of the ACGT infrastructure (in particular, tasks and challenges); (3) judgments regarding the project outcome and science policy; and (4) judgments regarding the anticipated profit of ACGT for cancer research and systems biology.

The interviews were digitally recorded, anonymized, and literally transcribed. Interviews with German participants were conducted in German; if cited, these interview passages were translated afterwards into English. The empirical results are based on qualitative content analysis by using the software MAXQDA 11. First, the interviews were paraphrased and sequenced. Then, we created headings (categories) for individual statements and compiled topically similar statements. This resulted in main headings characterizing the topics that were jointly discussed in the interviews (Meuser and Nagel 1991). Below, the interview citations are characterized by the professional background of the interviewee. Heuristically, the citations are used to describe facts, circumstances, and situations in a narrative and comprehensive language and to prove personal statements and judgments.

In addition to the interview material, we analyzed the content of internal ACGT documents accessed via the ACGT intranet (e.g., descriptions of work, progress reports, newsletter volumes, reviews, meeting minutes, deliverables, and conference presentations) and publications that were published by the interview partners. Internal documents reveal original goals, project progression, self-representation, and evaluation by external reviewers, whereas publications offer more background information regarding the research being done in the ACGT project.

#### ***4.1.2 Needs and Demands for Coordinating Systems-Oriented Research in Oncology***

In order to process and share data from heterogeneous data sources, ICT was applied to systems research and its fields of application. In the late 1990s and early 2000s, the first digital databases were established to support systematization and integration of data from research of a given domain, for example, a model organism or a

disease, in a formalized manner (Leonelli and Ankeny 2012, 30). In cancer research, one of the most prominent examples of such databases is the cancer Biomedical Informatics Grid (caBIG) launched by a US government program.<sup>4</sup> In 2004, the National Institutes of Health started implementing an open-source ICT infrastructure for data sharing among organizations by developing software tools, data-sharing policies, and common standards and vocabularies to facilitate data sharing.<sup>5</sup> The systems-oriented research community in cancer and further globalized scientific communities have appreciated such solutions for capturing and sharing data and information, in particular if the established digital databases allow and promote cooperation with other databases, thus providing a platform for community building. “There are clear pragmatic advantages to this form of digital technology, which include ease of access on a global basis, the ability to maintaining and update them dynamically and at relatively low cost, the ability to simultaneously access various types of information for comparison, the open access to all interested researchers, and so on” (Leonelli and Ankeny 2012, 31). Systems-oriented research in oncology especially benefits from ICT support, as one of the interviewees of our case study pointed out:

I think cancer research does have a higher degree of complexity. Because of the heterogeneity of cancer in general, because it is not a single disease, there are so many diseases that are completely molecularly and genetically different. And also the data that is being collected is very heterogeneous, very complex. It requires a lot of work to analyze it, to understand it, to use it. So I think that complexity was basically suggesting the need for [ICT] solutions. And also the fact that at some point people see the need to collaborate and to work together. So they do it on this multicentric trials, they want to share data, they want to analyze data together and so I think it is a very good domain, a very good model for trying to set up this multidisciplinary type of projects. (Interview [I] 2, IT)

In the context of clinical oncology, classical clinical trials are conducted in various phases, whereas each phase has a different objective involving different groups of patients, for example, medicalized group and randomized control group. Post-genomic clinical trials, however, cannot be conducted with current methodologies; they are characterized by the fact that molecular technologies are used—sometimes different kinds of molecular technologies in one clinical trial—and that large data sets are needed for statistical analysis. The need for statistically relevant data sets is challenging, because some biologically distinct patient groups may be represented only in small numbers. Hence, some predictions are limited inasmuch as robust classifiers that work well for predicting outcome in well-represented patient populations may, in fact, not work well in underrepresented groups. For example, one of the important prognostic biomarkers for breast cancer is the status of the estrogen receptor (ER). The majority of tumors examined within the good-prognosis group are ER-positive, but it is not yet clear whether the ER-negative group will not

---

<sup>4</sup>caBIG, <http://cabig.cancer.gov>. Accessed January 6, 2014.

<sup>5</sup>In 2011 a report on caBIG raised significant questions about effectiveness and oversight. As a consequence, its budget and scope were significantly trimmed. In May 2012, the National Cancer Informatics Program (NCIP) was created as caBIG's successor program. caBIG, <http://cbiit.nci.nih.gov/ncip/about-ncip>. Accessed February 21, 2014.

develop cancer later on and how well this predictor works in a larger cohort of ER-negative patients (van't Veer et al. 2002). In order to find this out, large patient cohorts representing the whole spectrum of a given cancer are needed for molecular profiling and statistical analysis. Such numbers are only accessible in multicentric international trials.

Today, the amount of data is increased by complex research designs with large patient cohorts. However, large-scale data sets are also due to the application of high-throughput technologies. For instance, one biological sample can be used to generate many kinds of data in parallel, such as genome sequence, patterns of gene expression, metabolite concentrations and fluxes, and so on. Furthermore, because of the continuous development of molecular technologies (e.g., next-generation sequencing methods), new types of data are continuously introduced into research. Generally, all the data are immediately integrated into data collections as databases and infrastructures are used to store as much data as possible (Ankeny and Leonelli 2011). In systems-oriented research in oncology, very heterogeneous data types such as Omics data, clinical data, imaging data, and pathology data are brought together and managed by ICT support.

Well, I think that the initial idea – if you go back to the vision of the project which was influenced by similar visions and initiatives that have been already established in the United State- was the fact that due to the developments of the time, the new types and size of data generated through developments in the biological domain, molecular biology, and the new types of technology generating tons of new types of data—proteomics and other types of data—we realized that the key problem was the fact that there were a lot of inefficiencies in the pipeline of trying to bring together diverse types of data, diverse tools of technology that need to exist in analyzing those data, and support more efficient ways of distributing teams that by nature are involved in such interdisciplinary types of research, clinicians like molecular biology, computer scientists, etc. So there are a lot of inefficiencies in the process of semantics, harmonizing the data, and the representation of the data, developing shared tools. Therefore, to support this concept of open source sharing of tools, avoiding reinventing the wheel, etc., since every specific lab invests in developing their own computational solutions and platforms. And therefore, the vision and the ultimate objective was to establish an infrastructure that would attempt to move forward toward a more efficient way of managing data, sharing data, sharing tools, and enabling distributed collaborators to work as a virtual type of an organization supported by an information technology solution. (17, IT)

The quotation indicates that the starting point to bring the ACGT project into being were not only new tasks and approaches in data management due to the high-throughput production of postgenomic data. The interviewee mentioned in the same breath that the consortium members have been aware of obstacles and inefficiencies in the analysis of such data as their variety and volume increase tremendously.

### ***4.1.3 The Development of ICT Infrastructures***

The vision and ultimate objective of the ACGT project was the development and realization of an ICT infrastructure that aims at meeting the needs and demands of postgenomic cancer research. Above all, ICT infrastructures have to facilitate

integrated access to heterogeneous data sources, for example, data from international distributed clinical trials, *in vitro* experiments, scientific literature, or Omic platforms.

Integration as a prerequisite of sharing data is often conceived of as the major problem or at least as a major challenge in systems biology (e.g., O'Malley and Soyer 2012, 61). In the following section, we take a closer look at the activities of facilitating integration of different data types and databases to build up an effective ICT infrastructure supporting systems medical research. Five challenges have been identified that have to be addressed to create an ICT infrastructure and, hence, to make data integration and sharing possible: (1) challenges of data acquisition, (2) ethical–legal challenges, (3) challenges of interdisciplinarity, (4) technological challenges, and (5) challenges of standardization. As discussed in the conclusion of this section, the analysis of the challenges points at the track of how the integration of data and, looking at the broader context, the understanding of biological systems is deeply shaped by ICT and their underlying design and conceptualization.

#### 4.1.3.1 Challenges of Data Acquisition

In order to investigate how disease-perturbed networks function, systems medicine needs different types of data generated from the patients (see Sect. 4.1). In addition to the different molecular data extracted from biomaterials, data are collected from clinical studies and health care. Clinical studies are usually done in clinical trials that are designed to answer specific questions about biomedical or behavioral interventions prospectively. Interventions to evaluate the effects on health outcomes, for example, include but are not restricted to drugs, cells and other biological products, surgical procedures, radiologic procedures, devices, behavioral treatments, process-of-care changes, or preventive care.<sup>6</sup>

Clinical data from health care are collected from patients who are diagnosed and treated in the clinic. Symptoms, diagnosis, histology, medications, treatment responses, and information related to lifestyle or environmental factors are stored on so-called patient records. According to the systems-oriented approach, it is furthermore claimed that the patient records ought to include the temporal dimension of biological parameters as well (Wolkenhauer et al. 2013, 503).

Hence, these three data types used in systems medical research are coming from different contexts and acquired for different purposes: molecular data are raised in laboratory research for investigating the molecular mechanisms of diseases; clinical study data are acquired in clinical trials for determining the safety and efficacy of interventions; and health care data are collected on a patient record for the purpose of reporting the patient's individual diagnosis and response to treatment regimes. The genesis of the data type has an impact on the validity and reliability of data as one of the laboratory researchers explained.

If you are using data only for research, you are doing research on a large number of patients to get some general conclusions. If you want to use it as a diagnostic tool, you want a conclusion for one individual patient at a time. It has to be much more precise, much more

---

<sup>6</sup>See World Health Organization, <http://www.who.int/ictrp/en/>. Accessed November 30, 2014.

reproducible, much more standardized. This is most of the time done by companies, which is really done in standardized labs and not research labs. (I10, BioMed)

From the quotation it follows that treatment decisions in clinical care must rely on individualized data that have to be precise, reproducible, and standardized. Research data, on the other hand, must serve statistical calculations and don't need to have individualized validity. Patient records contain only data of an individual patient; however, they also have a problem concerning data reliability and validity as one interviewee with a clinical background highlighted in the interview.

I'd rely on the data in the clinical information systems even less than on all kinds of other data. I would only rely on data that were gathered in clinical studies that were structured prospectively. Because they were defined in these case report forms from the outset. But all the unsorted stuff that's in the clinical information systems isn't structured at all. There are pathology reports, surgery reports, but it's all simply text. And then I need good data mining tools to dig out the required information. And if it isn't entered in a medical report in a structured way, then something might be missing, well, because whoever dictates the report may dictate two different reports for two patients with the same diagnosis. (I18, BioMed)<sup>7</sup>

As the interviewee stressed, he only assumes from prospective clinical trials that patient data collections are structured on a reliable base. However, even these data sets do not maintain reliability as long as data are not systematically managed and curated.

[O]ur problem really is the data. If you check the volume of the tumor. Or if I have this progression [of the tumor], if I want to be able to do something meaningful with the data, then they have to be good enough that you can rely on them. The data have to, even if I have an oncology patient and I start to gather data, then I have to know two years after I've gathered them, is he alive, or has he died? And if I don't follow that up, in other words, if I don't do data curation, then after ... what do I know, after a certain time, the data are useless. Well, because I'd get incorrect results again. That means, I have to involve the patient so that I get information about it, also about data curation etc., to get good results in the end. (I18, BioMed)<sup>8</sup>

---

<sup>7</sup>German original: "Auf die Daten, die in Klinik-Informationssystemen drin sind, würde ich mich noch weniger drauf verlassen als auf alle möglichen anderen Daten. Ich würde mich nur auf Daten verlassen, die in prospektiv strukturierten klinischen Studien erhoben worden sind. Weil die von vorn herein in diesen Case Report Forms definiert worden sind. Während das, was da lose in den Klinik-Informationssystemen drin steht, ist überhaupt nicht strukturiert. Da sind Pathologieberichte drin, OP-Berichte drin, aber das ist einfach Text. Und dann brauche ich gute Data Mining Tools, um daraus die nötigen Informationen rauszuholen. Und wenn ich das nicht strukturiert in einen Brief rein gebe, dann fehlt vielleicht auch etwas, ja, weil ich dann irgendwas diktiere oder derjenige als Arzt, der was diktiert, der diktiert bei dem einen Patienten, der die gleiche Diagnose hat wie der andere, zwei unterschiedliche Berichte." (I18, BioMed)

<sup>8</sup>German original: "[U]nser Problem sind wirklich die Daten. Wenn man guckt, wie groß ist das Tumolvolumen? Oder wenn ich diesen Verlauf [des Tumors] habe, damit ich richtig mit den Daten was anfangen kann, müssen die ja so sein, dass man sich auf die verlassen kann. Die Daten müssen, auch wenn ich einen onkologischen Patient habe und ich sammele jetzt Daten, dann muss ich aber nach zwei Jahren nach dem Sammeln immer noch wissen, lebt der oder ist der schon gestorben? Und wenn ich das nicht nachverfolge, also keine Data Curation mache, dann kann ich nach ... was weiß ich, nach einer gewissen Zeit mit den Daten auch nichts mehr anfangen. Ja, weil ich wieder falsche Ergebnisse kriege. Das heißt, ich muss den Patienten mit einbinden, dass ich darüber Informationen, auch über Data Curation kriege und etc., um dann am Ende auch gute Ergebnisse zu bekommen." (I18, BioMed)

The reliability and validity of data have an interdisciplinary dimension as data originate in different contexts and are usually gathered by scientists of different disciplines. The following citation shows that epistemic and pragmatic differences impede the understanding of how reliable data can be acquired in another scientific context.

The idea about which data clinicians can supply was completely abstruse in the beginning. [...] If I have patient data on a disease, then the scientist expects, okay, I've got my case report files here, and I've defined all the important things. And then each patient has exactly the same data set. And that is precisely not the case. Take a disease, for example a brain tumor. And then the question arises ... I'll give you two examples. The requirement I have to fulfill, I have to inform them now what the volume of the tumor is during the course of therapy. So I said, I can't do that. I can't tell you what the volume of the tumor is for glioblastoma. It's a highly malignant tumor. [...] And then I showed [them] images of the same patient in a T1, in a T1 with a contrast agent, T2 ... in four different modalities. The same plane through the skull. And all of them show the same diagnosis for the same tumor on the same day. But it looks different in each of the images. Because the different modalities show the tumor in different ways. And then there was the question: how big is the tumor in this image, how big in that image, how big in that one, then it turns out that it has four different sizes depending on which modality I use. And how big is it in reality? Nobody can answer that question. I could take the easy way out, a lot of people have done that, they deliver data, but they don't discuss the data with the people who want to use the data. So then I give them data where I say, the tumor had such and such a size at this point in time and such and such a size at that point in time. And then they do their calculations, and in the end everybody's surprised that the results are useless. Because the requirements for the data simply aren't right. And that's a very important point, that you convey the knowledge that you have in such a way that they understand: the data that we as clinicians can provide are biological data, they're completely different from mathematical data. (I18, BioMed)<sup>9</sup>

---

<sup>9</sup>German original: "Die Vorstellung, welche Daten die Kliniker liefern können, war am Anfang völlig abstrus. [...] Wenn ich hingehe und Daten von Patienten zu einer Krankheit habe, dann erwartet der Scientist okay, ich habe hier meine Case Report Files und habe definiert, was es alles gibt. Und dann hat jeder Patient genau den gleichen Datensatz. Und das ist eben nicht der Fall. Ich habe eine Krankheit, wie zum Beispiel einen Hirntumor. Und dann kommt die Frage auf ... Ich mache da zwei Beispiele. Die Anforderung an mich, ich habe denen jetzt mitzuteilen, wie das Tumolvolumen unter dem Verlauf der Therapie ist. Da habe ich gesagt, kann ich nicht. Ich kann euch nicht sagen, wie das Tumolvolumen beim Glioblastom ist. Also ein hoch maligner Tumor. [...] Und dann habe ich von dem gleichen Patienten Bilder gezeigt in einer T1, in einer T1 mit Kontrast, T2 ... in vier verschiedenen Modalitäten. Die gleiche Ebene durch den Schädel. Und alle zeigen am gleichen Tag bei Diagnose den gleichen Tumor. Aber der sieht in allen Bildern unterschiedlich aus. Weil die unterschiedlichen Modalitäten den Tumor unterschiedlich darstellen. Und dann war die Frage, wie groß ist der Tumor auf dem Bild, wie groß ist er auf dem Bild, wie groß ist er auf dem Bild, und dann kommt raus, der ist vier Mal unterschiedlich groß, je nachdem, welche Modalität ich benutze. Und wie groß ist er in der Realität? Das kann keiner beantworten. Ich könnte es mir einfach machen, das haben viele vorher gemacht, die liefern Daten, diskutieren die mit den Leuten aber nicht, die die Daten benutzen wollen. Dann liefere ich denen Daten, wo ich sage, der Tumor hat die und die Größe zu dem Zeitpunkt gehabt, hat die und die Größe zu dem Zeitpunkt gehabt. Und dann rechnen die, und am Ende wundert man sich, es kommt eigentlich nichts dabei heraus. Weil die Vorgaben der Daten einfach nicht stimmen. Und das ist ein ganz wesentlicher Punkt, dass man das Wissen, was man hat, auch so vermitteln kann, dass die verstehen, die Daten, die wir als Kliniker liefern können, sind biologische Daten, die sind völlig anders als mathematische Daten." (I18, BioMed)

To summarize, three different data types were identified in the interviews relevant to systems medical research. These three data types are acquired in different institutional and disciplinary settings (molecular research, clinical research, health care). Because of the different context of genesis, the data types used in systems medicine have different preconditions regarding data validity and fulfill different levels of reliability. Hence, according to data acquisition and management practices, the quality of data raised and stored for systems medical research can highly be variable. This is an important challenge when it comes to data integration: because of the interdisciplinary claim of systems medicine (and systems biology), a shared understanding of data reliability and validity in the respective acquisition context is obligatory before data can be integrated and related to each other.

#### 4.1.3.2 Ethical–Legal Challenges

The meaning of data not only depends on the scientific preconditions of the specific data acquisition context but also on relevant ethical–legal requirements. From this point of view, in particular patient records are severely restricted and managed differently than research data. Normally, a patient records are only accessible to the patient’s physician as well as a few other people directly involved in patient care who are obliged to maintain medical confidentiality. Clinical research conducted on patient data (and samples) is usually bound to an informed consent given for a single research purpose, that is, participating in a clinical trial, by the trial participant (Kollek 2009). Instead, molecular research data are normally authorized by a broad or blanket approval (Coebergh et al. 2006) to use the data for unlimited research purposes, such as the development or evaluation of new diagnostic tools, genetic studies, or biomarker identifications. As systems medical research aims at integrating clinical and molecular data, the different ethical–legal requirements of confidentiality and protection of data have to be served. The ethical–legal requirements for data management increase when it comes to postgenomic research in international research settings as one of the interviewees pointed out.

For example, what we like to see is that most of the trials they are being done by different hospitals. These hospitals can be in the same country, but most of the time they are also in different countries and so you have to make sure that legally everything is okay, in the framework for the trail across the countries, because different countries can have different laws. Then you have to make sure that everything is also legally and ethically fine for the sampling and the shipping of the samples, where the analyses are being done. (I10, BioMed)

To address the ethical–legal requirements for trans-European research projects, the ACGT infrastructure provides a data protection framework, which was designed as a safety net consisting of three pillars (Forgó et al. 2010, 102). The first pillar was the development of a network of trust within the project. A legal body taking over the responsibility of the data controller (data protection authority), the involvement of an internal security authority (trusted third party), the conclusion of legally binding contracts and, finally, the ICT security tool called the Custodix Anonymisation Tool (CAT) ensured that the data would be estimated as de facto anonymous data

within the research network. The second pillar ensured patient involvement. This was achieved by obligatory patient information and the requirement of informed consent as well as the establishment of a central contact point for all participating patients. The third pillar was the identification of provisions that allow the processing of personal data for research purposes. This was done on the basis of a thorough analysis of different national legislations. Taking the three pillars together, scientists can use the ACGT platform knowing that the use of data is ensured by the data protection framework.

The goal of the data protection framework is to come to an environment where whenever you add new applications, new tools, you as an end user, as a partner, or as a patient would know that by default it would be in compliance with all laws and it would stick to the same ethics code that governs our platform. So basically you would know that it would be safe. Compliance by default that is what we want to reach. So if you plug in, and you are allowed to plug in, then everyone around you knows it must be okay. That is the achievement that we wanted to reach. (I11, IT)

In conclusion, data are scientifically as well as ethically–legally embedded into research traditions and international standards and guidelines. The ACGT consortium was very aware that different ethical–legal requirements of molecular research, clinical research, and health care have to be served to gather data from patients in the different settings. The ethical–legal framework, translated into legal contracts between the different stakeholders participating in clinico-genomic research on cancer, was fully developed after the completion of the ACGT project and is currently used in ACGT follow-up projects.

#### **4.1.3.3 Challenges of Interdisciplinarity**

Systems biology and, even more, systems medicine is a genuine interdisciplinary field targeting the modeling, understanding, and finally manipulation of living systems. From the very beginning, proponents of the new systems approach stress that systems biology must and will be able to display a holistic view on processes of life explaining how cells, tissue, and organisms interact from the workings on the molecular level (see Sect. 2.5). Hence, systems biology brings together scientists from a variety of disciplines such as mathematics, computer sciences, medicine, and biology. The interdisciplinary claim of systems-oriented approaches is, however, another prominent challenge in the development of systems biology and systems medicine as different scientific research logics, theories, methods, practices, and discourses come into play and interact. In the ACGT project, computer scientists worked together with biologists, clinicians, and experts from law and ethics. Asked for interdisciplinary problems in the interviews, the scientists often referred to misunderstandings in the communication across disciplines.

But I think what we have to say is that, for me honestly, it took me six months to one year before all the partners could communicate a bit in their projects, because the backgrounds were so different that it takes some time to understand the vocabulary and the backgrounds of everyone before you can move together. (I10, BioMed)

Well, it is difficult to talk with each other across different expertise fields. Because people use different technologies, people have very different objectives and so what you need in

order to be able to communicate is a form of respect. And it is often difficult to create respect across different scientific fields or technological fields. You know, because the one feels superior to the other. That is something that you see a lot. (I11, IT)

The interviewees stressed that interdisciplinary communication is built on the acknowledgment of different backgrounds and the comprehension of disciplinary vocabulary as well as respect to listen to other disciplines. Several interviewees mentioned that it finally took up to a year to bridge the language issue and to understand and work with reference to each other.

For me, the first year was the decisive year in which people had to learn to agree on a language so that everybody understood what was meant. And that's also something that reflects all the semantic integration in the project. If I want to use the data, then I have to generate the data in such a way that they can really be combined with one another in a simple way. And for me, in the beginning, in the first year, that was pretty—how should I say—where I thought, I don't understand a thing here. (I18, BioMed)<sup>10</sup>

Another interviewee explained interdisciplinary differences by referring to different disciplinary modes of thoughts. His example was the ACGT Master Ontology that was built for clinicians but did not, in his view, stand a chance in the clinic because of the differing mindsets of the ICT experts who had created the ontology and the clinicians who wanted to use it in daily work.<sup>11</sup>

For a clinician, the ontology is cumbersome to use. This thing hardly has a chance of becoming widely used in clinical work, because the way of thinking of a person who develops an ontology is entirely different from how a clinician thinks. That's why in ACGT, there was already the idea to develop a tool that represents the ontology in a way that the clinician can understand it. The ontology is structured like a tree. And the clinician's thinking may also be like a tree, but he thinks in a way where the patient comes first, he wants to have the diagnosis first of all, then there are diagnostic measures, then there are therapeutic measures and so on and so forth. That means, this tree is represented in a completely different way than how the ontologist represents such a tree. And that's a clinical view of an ontology. And it's absolutely difficult to develop it. ACGT didn't make a lot of progress there. (I18, BioMed)<sup>12</sup>

---

<sup>10</sup>German original: "Das erste Jahr war für mich das entscheidende Jahr, in dem die Leute lernen mussten sich auf eine Sprache zu verständigen, so dass jeder verstand, was damit gemeint ist. Und das ist auch etwas, was ja die ganze semantische Integration in dem Projekt widerspiegelt. Wenn ich die Daten verwenden will, muss ich auch die Daten so machen, dass die sich untereinander wirklich einfach verbinden lassen. Und das war für mich so am Anfang, im ersten Jahr, ziemlich—wie soll ich sagen—wo ich dachte, ich verstehe hier überhaupt nichts." (I18, BioMed)

<sup>11</sup>The ontology formally represents knowledge as a set of concepts within a domain (such as a disease), and the relationships between those concepts (see the following Sect. 4.1.3.4).

<sup>12</sup>German original: "Die Ontologie ist für die Benutzung für einen Kliniker schwerfällig. Das ist ein Ding, das kaum eine Chance hat, in die Klinik einzuziehen, weil die Denkweise eines Menschen, der eine Ontologie entwickelt, völlig anders ist als ein Kliniker denkt. Deswegen war in ACGT schon die Vorstellung, man muss hier ein Tool entwickeln, was die Ontologie abbildet in einer Art und Weise, dass der Kliniker das versteht. Die Ontologie ist ja so baumförmig aufgebaut. Und der Kliniker denkt vielleicht auch baumförmig, aber der denkt in einer Weise, wo am Anfang der Patient kommt, der will die Diagnose erst mal haben, dann gibt es für ihn diagnostische Maßnahmen, dann gibt es therapeutische Maßnahmen und so weiter, und so fort. Das heißt, dieser Baum ist völlig anders abgebildet als der Ontologe so einen Baum abbildet. Und das ist ein Clinical View auf eine Ontologie. Und die zu entwickeln, ist absolut schwierig. Da ist ACGT nicht sehr viel weiter gekommen." (I18, BioMed)

Taking the quotations together, it becomes evident that interdisciplinary collaboration is always challenging, even if the scientists already have experience in other interdisciplinary settings. The ACGT interviewees described forthrightly that it took them the first year of the project to align terminology, mindsets, concepts, and expectations before the joint research even began. Maybe for that reason, they stress that progress and success of the ACGT project was mainly grounded in overcoming the interdisciplinary challenges and initial misunderstandings between the disciplines involved. However, the ontology was given as an example that in particular, differences affecting the mode of thought and research logic sometimes persist. This example shows that ICT as applied science in the interdisciplinary setting sometimes restricts the articulation of theories and the development of a research logic specific for systems biology.

#### 4.1.3.4 Technological Challenges

One of the most prominent challenges frequently brought up in our interviews was technological problems regarding storing, integrating, and accessing high-throughput data. Classical approaches to solve these problems focus on syntactic interoperability which means that two or more databases have to be capable of communicating and exchanging data (Sujanski 2001). Technically, this requires a software component called a parser that analyzes input data to build the underlying data structure. The structural representation of the input, often described as an abstract syntax tree or other hierarchical structure, facilitates that different data and message formats (e.g., data-exchange protocols, programming languages) can be interconnected to an application programming interface called the data abstraction layer. Finally, the data abstraction layer is able to unify the communication between a computer application and databases by representing the data structures in a unified data and message format such as a programming language, for example, the eXtensible Markup Language (XML) or Structured Query Language (SQL).

Another critical feature to create interoperability between different data sources refers to the meaningful and accurate interpretation of the information exchanged. Here, the absence of shared terminology is one of the basic obstacles to enable communication and sharing of data (Tsiknakis et al. 2006, 248; Burgoon 2007, 404). Semantic uncertainties often refer to conflicting terminologies and classifications, or in other words, to missing agreements on terms and concepts. A basic tool to homogenize terminology and to build semantic interconnections is the ontology (Rubin et al. 2008). The ontology formalizes the meaning of terms through a set of assertions and rules that are collectively known as description logics. The ontology is concerned with what concepts are contained within the field, what information is required for each concept to have existence, and how different concepts are related to each other. Therefore, it depicts concepts within a domain (such as a disease), and the relationships between those concepts.<sup>13</sup> There is no need to attach any language

---

<sup>13</sup> Philosophically, ontology is a part of metaphysics dealing with questions concerning the nature of being, becoming, existence, or reality and their relations (e.g., Burkhardt and Smith 1991).

term to the classes as the ontology can be built in a language-neutral way. However, it can be done, inasmuch as naming the classes fosters the ontology's transparency to the users. For them, ontologies offer a structured knowledge repository which is used to describe the domain and can be used to reason about the entities within that domain. Bio-ontologies are already acknowledged as a relevant method for database integration in systems biology (Wierling et al. 2007). The Gene Ontology, for example, has been continuously developed since the late 1990s to classify, exchange, and compare data about gene products of a wide variety of species (Leonelli et al. 2011).

The ACGT consortium did not adapt to an already existing ontology, but decided to create an ontology that met the specific needs of the ACGT project.

We studied and came up with solutions to [...] achieving this semantic data integration based on some specific and strange assumptions, which are the following two: that data reside and are under the control and responsibility of the data producing entity, so that data belong to the lab and the group and the individuals that are responsible for producing them. And the second hypothesis we build our solution upon is that this integration is achieved through the use of a shared conceptualization of what we call the master ontology and there are appropriate techniques for utilizing this master ontology in achieving integration at the level of data. (I7, IT)

The ACGT Master Ontology (ACGT MO) was hand-tailored for the use in post-genomic cancer research. In particular, it structures and describes the concepts that are important in the domain of postgenomic clinical trials on nephroblastoma and breast cancer (Brochhausen et al. 2011). As the data stay in the original databases, a translation or mapping of the data is necessary to interlink the data to the ACGT platform. "We needed a mapping that says, well, within the project this is how we refer to a patient, this is how we refer to a microarray dataset, this is how we refer to ... So there was an agreed list of terms and then the semantic mediator has to do the mapping; so it says, well, this entry here it actually maps to this term in a global dictionary." (I1, IT)

Within ACGT, the semantic mediator is a software tool that harmonizes data contents to make heterogeneous data acquirable to the components of the ICT system, "So that data are more than just bits and bytes, so that the other parts of the system can understand them" (I1, IT). Technically, the semantic mediator systematically coordinates data from different data sources by performing query translations from the ACGT MO to the local databases.

Being able to computationally—because an individual will not look into a contextual description of what this algorithm does etc.—being able to computationally assess the capabilities of a specific tool, a specific service, you need to describe these capabilities: inputs, outputs, types definitions, etc. of data in a way that makes sense to a computing system. That is the essence of metadata. So in a sense rather than having the producer of this tool describing in a page or half a page verbally through text, that is algorithm. It models a specific function etc. and requires a set of input. You need to do that through elements that describe this capabilities and requirements. That is what we call metadata. (I7, IT)

Basically, the metadata help to categorize the data coming from different data sources to map or define the data for further data processing. Hence, each data type

has a profile defining how the data have to be treated. Once the profile is in place, all data of the same type are processed in the same way. In this respect, the generality of the ACGT MO causes problems, because it intends to represent the whole domain of cancer, whereas the databases are normally developed with a specific goal in mind. As a consequence, when trying to explore a database by formulating queries using the ACGT MO as a guide, it is likely that the query term cannot be found in the specific database (Bucur et al. 2011, 1124).<sup>14</sup>

Within the ACGT MO, the user can search for terms or concepts as in a dictionary. In addition, the ontology viewer visualizes the interrelations between the concepts as a tree-like structure of the ACGT MO. In this context, it is important to notice that many existing ontologies focus on the classes or categories of the entities in a given domain. These ontologies might give a hierarchy of those entities via the basic taxonomical relation, the *is\_a* relation. But only the inclusion of additional semantic relations between classes, for example, *x is part of y*, *z is adjacent to u*, *a is prior to b*, can lead to a comprehensive representation of the phenomena (Brochhausen and Blobel 2011). Taking this into consideration, users of the ACGT MO can create their own semantic tree by setting different kinds of semantic relations that are useful for their scientific observations (Brochhausen et al. 2011).

The ACGT MO finally contained more than 1,600 classes and nearly 300 properties. The ACGT consortium applied it to the Open Biomedical Ontology Foundry, which is an open source initiative to create a suite of orthogonal interoperable reference ontologies in the biomedical domain.<sup>15</sup> However, the ACGT MO didn't succeed in going through the quality assurance of the OBO Foundry until the end of the project, which was the last step before actually becoming an agreed ontological element of the OBO Foundry in the domain of cancer.

To conclude, the development of the ACGT MO itself was a complex and challenging task. Unsurprisingly, one of the interviewees made very clear that semantic interoperability has a much higher level of complexity than the syntactic interoperability: "The semantics is more difficult than the syntax, because understanding the syntax of something or agreeing on the syntax doesn't guarantee that you know the meaning. And mostly the meaning is harder to agree upon." (I2, IT) The ACGT consortium tried to tackle the considerable challenges such as the specification for the domain of cancer or the mapping of flexible interrelations between the different terms and concepts. However, its incomplete process of application to OBO Foundry indicates that semantic standardization is a long-lasting endeavor that needs further developmental steps with regard to approval and sustainability within research communities. Even if foundational semantic work, in particular in the domain of metadata definitions, was done in the course of ACGT, from today's perspective it seems too ambitious to set up an ontology from scratch to be used in clinical practice in time-limited research projects.

---

<sup>14</sup>ACGT MO is written in the Web ontology language OWL-DL. This ensures that all conclusions drawn from the ontology are guaranteed to be computable. For example, OWL-DL allows automatic reasoning (e.g., for consistency checking).

<sup>15</sup>OBO Foundry, [www.obofoundry.org/](http://www.obofoundry.org/). Accessed September 19, 2014.

### 4.1.3.5 Challenges of Standardization

The previous section showed that the ACGT consortium tackled syntactic and semantic integration of data and tools technologically. This approach usually includes the attempt to standardize such technologically driven integration processes. With regard to ICT infrastructures, standards ensure the line of communication between data and tools as well as tools and users (Hanseth et al. 1996, 410). However, the molecular technologies used to generate data for systems biological research are challenging the communication processes as these technologies are fast-evolving and, hence, may alter data that are supposed to be stored in standardized databases or infrastructures.

And I think that most of the technologies are so immature and fast-evolving, therefore, it's practically impossible to follow this kind of development. So there are so many new data, so many new technologies everywhere. They are appearing on the landscape, there are so many of them disappearing from the labs or from the practice. So in some sense this kind of postgenomic research is hard to harmonize. (I6, BioMed)

Cutting-edge technologies keep the daily working processes in motion, in particular as reliable standards are often missing. As a result, the researchers are confronted with an almost too confusing amount of different technologies and standards referring to these technologies.

As I already said there are lots of standards that evolve. There are standards at the lowest level of the IT, for example, Web services, the whole network exchange, how do computers exchange messages. One of the things of ACGT was the distributed system, so there were services and resources distributed all across Europe. We also had the Grid infrastructure and the Grid services set their own standards. They are still evolving and there are on top of that the genetic standards of how you express sequencing information and the clinical standards and the query language standards. If you look at all the standards that were evolved, you tend to get more and more standards that are not relevant. (I1, IT)

However, many scholars refer to the urgent need for standards to describe, format, submit, and exchange data (e.g., Green and Wolkenhauer 2012, 769). The short and insecure innovation cycles of high-throughput technologies and the interdisciplinary approach in systems medicine increase the demand for standards that are reliable and accepted across the community of users. But the technological and interdisciplinary innovations trigger multiple standard operating procedures at the same time (Auffray et al. 2009, 2). Thus, a growing number of unnecessary overlaps and duplications of standardization procedures evolves (Field et al. 2009, 234), in particular when ICT-based standards are involved as the following quotation shows.

There is almost a standard for anything, even worse. There are more standards for the same thing in many cases. So the problem is that if you are writing programs and you want to conform to a standard, you have to understand the way that you can apply this to your own software. And so like I said, it can mean that you download a piece of software from somewhere that conforms to the standard and you interface it. But then you have to understand this API that allows you to use it. That is basically the story that I was just telling. So for many people who do not have this understanding, it is just a matter of being pragmatic. You can either spend, let's say, a week of time to try and understand how to interface something that is not yours into your own software or you can just create something for yourself in an

hour. And so everybody chooses the last option unless you are a computer scientist. If you are a computer scientist it takes you one hour to conform to the standard. (I3, IT)

Of course, anyone can claim to develop a new standard but standards necessarily need approval as one of the interviewees explained in relation to formal and de facto standard-setting procedures. Top-down standardization is initiated by standard development organizations (e.g., ISO, SEN, HL7). They are usually entitled to develop formal standards for a specific setting. The other process is the bottom-up approach, where user communities or industry trigger de facto standardization.

For me, DICOM is the indicative: the digital imaging standard, which was an effort through the American Society of Radiologists and NEMA, the National Engineering Manufacturing Association. So the industry and the users jointly developed a standard that was pushed to become a universal standard. That is very difficult to happen actually in the context of a three or even four year EU funded project. You need for exploitation, for exploiting the results of the research project, you need structures that will leave after the end of the project. (I7, IT)

As the quotation indicates, neither formal nor de facto standards were developed in the course of the ACGT project. Missing sustainable structures for exploitation of research results are one aspect of hesitant commitment to formal standard-setting procedures, in particular in time-limited research projects. Another aspect is the dynamics of the research field and the individual interests of researchers, as another interviewee pointed out.

In a fast evolving field of research, the problem is that standardization is causing delays. In order to be compatible with other, in order to maintain this kind of compatibility, you have to slow down and put some effort toward this kind of end. However, it appears that the forefront, the people who are really on the edge of developments, neglects standardization and then move on. In that sense, whenever standardization has to be done, it has to be done by, let's say, the second line of, or the second front of research that is not that ambitious, but it may be as important. So I would agree that it is important, but it appears that nobody really who has ambitious scientific questions would spend time to serve the purposes of standardization. (I6, BioMed)

However, even if scientists are usually not interested in taking responsibility for standardization processes, they need to find an agreement on the standards that ought to be used in a research project. One of the interviewees described the daily working experience of how standards are chosen in research collaborations. Commonly, joint discussions take place on how standards should be set and what should be used according to the specific research purposes.

Often there are fifty per cent split by discussions on standards on which one is the best standard. From standards in data collection to standards in normalization processing of data. I mean there is a long list and in some cases, there are different standards that do different things. And so we have to decide on the standard that is best for the specific use of the data. It is a very complex question on how to ... and probably in most cases it is not good to just use one standard, because different ways of processing your data do give you different data. And they are better in some contexts rather than others. So it's a good question. And I don't have a general answer. We don't have a general process for using the standards. It's on a case basis. Of course, it is based on what other people are doing, what other census are doing. In some cases, it is not us choosing the standards, because if you are

part, like often we are, of a large international study, the standards are chosen together. (I9, BioMed)

In the course of the ACGT project, the approach was to first look and review what kinds of standards already exist and to reuse as many of them as possible. Usually, those standards are preferred that are supported by large communities of users. It is assumed that those ICT tools and services that are built on broadly accepted standards will be recognized and reused in the respected research community as well. Another interviewee precisely described the problems that occurred with picking standards for a particular task in the ACGT project, namely building tools for knowledge discovery workflows.

So, we had the question about standards regarding workflows, for example. How do I represent a workflow, how do I save it? Which data storage device do I use for the individual services, for example, how do I describe it? How do I describe whether it has a particular quality? How do I describe who built it, how can I verify if it's still okay? Well, the number of standards that you can use or that you could wish for is relatively high. The problem is just that there are simply very many of them. So, there's no single standard in the sense that really everyone uses it, but there's simply an incredible number of things that an incredible number of people have done in those areas, and where in the end, everyone picks out whatever they happen to need. That means, the only thing that really is a standard is if you decide to take a particular tool and then simply use the format that that tool uses as a standard. (I13, IT)<sup>16</sup>

As the interviewee precisely described it, in the process of creating new tools the ICT formats of existing tools work as standards. Therefore, newly developed tools such as the mentioned data discovery tool for workflow building can be directly linked to the chosen ICT format, which was in this case the programming language R for statistical calculations and graphs. Another good example for standardization based on ICT support is the so-called MIAME Convention describing the minimum information about a microarray experiment that must be provided to report data in microarray-based publications (Brazma et al. 2001).<sup>17</sup> According to the convention, the raw data have to be defined as data files produced by the microarray image analysis software. Even if the formats, annotations, and protocols are not prescribed, the convention includes a list of possible MIAME compliant software. This makes obvious that study designs to be approved have to be based on ICT. Again, ICT

---

<sup>16</sup>German original: "Also wir hatten die Frage nach Standards zum Beispiel bei Workflows. Wie stelle ich so einen Workflow dar, wie speichere ich den ab? Welchen Datenspeicher nehme ich bei dem einzelnen Services zum Beispiel, wie beschreibe ich den? Wie beschreibe ich, ob der eine bestimmte Qualität hat? Wie beschreibe ich, wer den gebaut hat, wie kann ich dann überprüfen, ob der auch noch in Ordnung ist? Da gibt es dann auch relativ viele Standards, die man brauchen kann oder die man sich wünschen könnte. Das Problem ist halt, dass es da einfach sehr viele gibt. Also es gibt keinen Standard in dem Sinne, dass wirklich alle ihn benutzen, sondern es gibt einfach unheimlich viele Sachen, die unheimlich viele Leute gemacht haben in den Bereichen, und ja, wo sich dann im Endeffekt jeder selber das zusammensucht, was er gerade braucht. Das heißt, das Einzige, was halt wirklich ein Standard ist, ist wenn man sich entscheidet, ein bestimmtes Tool zu nehmen und dann einfach das Format, das dieses Tool nimmt, als Standard zu nehmen." (I13, IT)

<sup>17</sup>See also [www.fged.org/projects/miame/](http://www.fged.org/projects/miame/). Accessed June 12, 2014.

software works as a standard enabling the unambiguous interpretation of the results of the experiment. The MIAME standardization process set up by the Functional Genomics Data Society (FGDS) was very powerful in the systems biological community. One possible reason was that gene expression was successful right from the beginning in the emerging fields of systems-oriented research and many microarrays were done. Taken the aspects outlined together, standardization procedures in systems biology are triggered by the need to make data and study designs comparable to integrate and share data and finally study results. As data storing and processing are based on ICT systems, ICT also deploys the standards for data quality, annotation, and exchange.

#### **4.1.4 Concluding Remarks**

In this section, we looked at the challenges of facilitating integration of different data types and databases to build up an effective ICT infrastructure supporting systems medical research. In the interviews, we identified at least five challenges relevant to data integration realized in an ICT environment.

The challenge that prepares the ground is data acquisition which takes place in different contexts to gather the three basic data types (molecular data, clinical study data, and health care data) to be used in systems-oriented research. As shown, the preconditions of data acquisition in the laboratory and the clinic are very different. Gathering molecular research data means in the first instance to hook up with new acquisition technologies and to deal with the scale and breadth of Omics data. Gathering clinical study data means following a prospective research protocol and to set up a recruitment process including informed consent. Gathering health care data stemming from treated patients means dealing with unstructured and incomplete data sets and with high data protection standards because of medical confidentiality. The interviewees broadly discussed that the different preconditions of the data acquisition challenge data reliability and validity. Hence, it was put forward that due to the data acquisition context, reliability of the different data types is highly variable. Furthermore, the ethical–legal requirements of the data types differ according to the data acquisition context. Generally, data access is only permitted by informed consent. However, the broad consent for research on biomaterial is usually given one time for unlimited use, whereas research conducted on clinical study data is usually bound to an informed consent given for a single research purpose. Access to health care data is most restricted as these data are primarily raised for treatment decisions and not for research purposes.

Of course, restricted access to medical data for research purposes has been broadly discussed in ethical debates. However, according to the interdisciplinary approach in systems medicine, data are no longer exclusively analyzed by the scientists who had raised them. Therefore, it is necessary that the nature of the data type (e.g., reliability, ethical–legal requirements) is known beyond its specific context. The interviewees generally acknowledged that interdisciplinarity is an ambitious

challenge, as misunderstandings between the participating disciplines often occur because of ignorance and differences in terminology, research practices, and logics. The interviewees stressed that progress and success of the ACGT project was mainly grounded in overcoming the interdisciplinary challenges. However, the interviews revealed at the same time that in particular, differences affecting the mode of thought and research logics sometimes persist.

The most prominent challenge for systems-oriented research is, however, the integration of the different data types coming from different data sources (e.g., O'Malley and Soyer 2012). The ACGT consortium tackled the challenge by setting up an ICT infrastructure addressing two levels of integration. In brief, syntactic integration provides the technological rack to facilitate data exchange, whereas semantic integration ensures that the data exchanged are accurately interpreted. A range of technological tools and services, such as the workflow builder or the ontology discussed earlier in Sect. 4.1.3, was developed to assist syntactic and semantic data integration. These tools and services finally prove that the challenge of data integration is exclusively approached by ICT-driven technology. This approach usually includes the challenge to standardize tools and services triggering the integration process. As the interviewees described it, new tools are directly built on ICT as the ICT formats of existing tools work as standards for the new ones.

Because of the nature and amount of data generated by different technologies, laboratories, researchers, and clinicians, data integration has triggered the implementation of multiple standard operating procedures based on ICT. In the face of the different scopes and formats, and different disciplinary origins and developments, standardization processes seem to be the central mechanism of coping with the overarching tasks of data integration for building up data collections. Not surprisingly, the data that are most successfully assembled into large data sets are genomic data which are produced through highly standardized technologies such as genome sequences or microarrays (Leonelli 2014, 5).

ICT-based standards and guidelines define what counts as reliable evidence, clear nomenclature, and commonly accepted experimental practice within the emerging field of systems medicine. This already has a sustainable impact on the handling of data: because of the computational environment, data are split into the pure data content and the data structure describing the content. Of course, the distinction of data and metadata is a phenomenon with a long tradition in biology (Edwards et al. 2011; Leonelli 2010). What is different in systems medicine (and biology) is the fact that the latter is first and foremost defined and attached by ICT. It finally forms a new body of information including formal data and message format, accurate classification, or other relevant ICT metadata. In this context, Sabina Leonelli (2014, 6) points out that the task to create the information of data classification (e.g., adding keywords, metadata, etc.) is usually one of the tasks of ICT-trained curators who have therefore gained influence on the meaning and interpretation of data in systems biological research.

To conclude, it is often stressed that the application of high-throughput technologies has made it possible to increase dramatically the amount of information that can be stored and integrated (e.g., Leonelli 2012). This assumption was verified by

the experience of the interviewees. However, our case study also revealed that this quantitative shift has brought the need to standardize data for data integration and reuse. We argue that the quantitative shift has led to qualitative changes of how to handle and use large repositories of standardized data in systems medical research: the significance and meaning of data have changed by defining which part is for scientific use and which contains more or less purely technical information such as the message format. Finally, data produced by almost fully automated and highly standardized procedures are rather regarded as a computer output reaching value because of the reproducibility and reliability based on ICT (García-Sancho 2012, 26). This has resulted in the acknowledgement of data “as key scientific components, outputs in their own right” (Leonelli 2014, 9) that need to be widely disseminated. Hence, ICT environments collect and process not only data; they construct at the same time the data by assigning significance, meaning, and finally, evaluation to data and its parts. From this it follows that understanding and modeling of biological systems are deeply shaped by ICT. In the next section we look in more detail at the process of how ICT and their underlying design and conceptualization shape the modeling of cancer in silico.

## 4.2 Simulating Cancer In Silico: The Oncosimulator

The goal of systems biology is to model molecules, cells, tissue, organs, body systems, and whole organisms holistically. One possible access to reach this ambitious goal is to study diseases and the alterations between normal and diseased biospecimens. Models to understand and predict the genesis and development of a disease such as cancer can not only be proved in vivo and in vitro, but also be theoretically analyzed with in silico techniques. In silico as a term describes the modeling, simulation, and visualization of biological and medical processes in computers referring to any application of computer-based technologies (Michelson et al. 2006). The increasing volume of molecular data and the decreasing costs of computational power have made it possible to run more and more in silico simulations today (Deisboeck et al. 2009).

For instance, the virtual self-surviving cell modeled by Masaru Tomita is regarded as one of the first whole cell in silico models (Tomita 2001). This modeled cell consists of 127 in silico genes, 120 coming from *M. genitalium* and 7 coming from other microorganisms. Based on the model, it was investigated how the alterations of glucose supply and signal pathways control the knockout of distinct genes. Observing the behavior of in silico cells yields comparative insights and can lead to the discovery of causalities and interdependences by providing in silico experimental devices for hypothesis testing and predictions (Gramelsberger 2013, 157).

In the clinical context, in silico modeling might finally lead to reliable predictions as to which treatment will fail in a patient before it is applied (Graf et al. 2009, 142). Thus, in silico oncology is one of the most visionary endeavors of the ACGT project concerning the actual use of systems approaches to medical decision making. We

therefore have chosen *in silico* technology as an example to analyze the conceptual development of tools and services for doing systems medicine. By looking at the development of an individual tool from scratch we aim in particular at investigating how ICT and their conceptualization shape systems medical research. The example under study is the oncosimulator incorporated into the ACGT infrastructure as an experimental platform. It simulates *in vivo* response of tumors and normal tissue to therapies based on clinical, imaging, histopathologic, and molecular data of a given cancer patient. In the long run, it aims at a better understanding of cancer at the molecular, cellular, organ, and body level and optimizing therapeutic interventions on a patient-individualized basis by performing *in silico* experiments of candidate therapeutic schemes.<sup>18</sup>

In the course of the ACGT project, Georgios Stamatakos and the *In Silico* Oncology Group at the Institute of Communication and Computer Systems, National Technical University of Athens<sup>19</sup> developed the initial version of the oncosimulator focusing on pediatric nephroblastoma, a childhood cancer of the kidneys, and in particular on a trial run by the International Society of Pediatric Oncology (SIOP) in collaboration with the Department of Pediatric Hematology and Oncology at the University Hospital of Saarland (Germany) led by Norbert Graf. For the first time, Stamatakos and his team were able to use real data before and after chemotherapeutic treatment (Stamatakos et al. 2007). This was a breakthrough to adapt the software to real clinical conditions and, at the same time, validate the software using real-world results. By using real medical data concerning nephroblastoma for a number of patients in conjunction with model parameters based on literature research, the tumor volume shrinkage has been predicted with reasonable accuracy. Up to now, the oncosimulator has been advanced and implemented in further research projects in the context of the 7th EU Framework Program such as the projects p-medicine and CHIC.<sup>20</sup>

### ***4.2.1 Vision and Definition of the Oncosimulator***

The genesis and progression of cancer is associated with tumor morphology, invasion, and related molecular phenomena (Sanga et al. 2007, 120). One of the grant challenges of the understanding of cancer progression is therefore to find the links between alterations and the hallmarks of cancer such as increased proliferation

---

<sup>18</sup>This work has been supported in part by the European Commission under the projects “ACGT: Advancing clinico-genomic trials on cancer” (FP6-2005-IST-026996), and “CHIC: Computational horizons in cancer: Developing meta- and hyper-multiscale models and repositories for *in silico* oncology” (FP7-ICT-2011-9-600841).

<sup>19</sup>*In Silico* Oncology Group, Institute of Communication and Computer Systems, National Technical University of Athens, <http://in-silico-oncology.iccs.ntua.gr/english/index.php>. Accessed May 3, 2014.

<sup>20</sup>p-medicine, [www.p-medicine.eu](http://www.p-medicine.eu) and CHIC, <http://chic-vph.eu>. Both accessed May 10, 2014.

and survival, aggressive invasion and metastasis, evasion of cell death, and increased metabolism (Hanahan and Weinberg 2011). However, it has been difficult to quantify the relative effect of these links on disease progression and prognosis using conventional clinical and experimental methods and observations. For example, the primary role of angiogenesis in promoting tumor growth and invasion has been well demonstrated, whereas the results of clinical trials using drugs to suppress neovascularization have not yet yielded unambiguous results (Kuiper et al. 1998; Bernsen and van der Kogel 1999). Hence, what is needed is a method to enable prediction of tumor growth and therapy outcome through quantification of the relation between the underlying dynamics and morphological characteristics.

The fundamental assumption underlying this approach is that any biological processes are amenable to mathematic and/or algorithmic description. In this regard, the genesis and development of cancer is regarded as a disease and at the same time as a natural phenomenon. From the cancer treatment perspective, what really matters is the discrete number of the usually few tumor cells surviving treatment and their discrete mitotic status (e.g., stem cells, cells of various mitotic potential levels, differentiated cells). Therefore, such a mathematical approach must apparently take into account both the deterministic and the stochastic character of the disease (Stamatakis et al. 2007). This challenge is tackled by a multidisciplinary method integrating mathematical description and computational simulation of the multi-scale biological mechanisms that constitute the phenomenon of cancer and its response to therapeutic regimes. By primarily applying discrete mathematics, the In Silico Oncology Group developed the modeling method called Discrete Event-Based Cancer Simulation Technique (DEBCaST) (Stamatakis 2011, 408).

DEBCaST is basically a top-down approach using clinical observations, including anatomic and metabolic tomographic images of the tumor, and the knowledge about the behavior of a cancer as a whole based on available physiological and biological findings. This information is required to identify subsystems of the tumor and to build a reproducible model of a specific cancer and its progression. Given that the discrete entities and quantities of a specific cancer in conjunction with their complex interdependences give rise to tumor relapse or ensure tumor control over a given time interval, the constant alignment with the clinical observations is required.

Multi-scale cancer models, which lie at the heart of the oncosimulator, should be driven by real clinical trials. This is completely different from the standard bottom-up approach adopted by most cancer modelers. Developing a tumor model by trying to exploit what you can do and by trying to extend what you can do is of course interesting and potentially useful. But the reality itself, at least the clinical reality, expects the modeler to be adapted as much as possible to the real clinical questions, the real clinical problems as they are posed within the clinical walls, let's say, or the clinical theatres. In that sense, a top-down approach rather than a bottom-up approach seems to be better in order to address such complex problems. This is my personal approach. [...] Anyway, models should be adapted and validated to real clinical trial data. The oncosimulator should undergo both retrospective and prospective clinical validation as a prerequisite to be translated into clinical practice. Models should be modular and extensible so as to be able to integrate new advances in cancer biology and clinical experience. (GS)<sup>21</sup>

---

<sup>21</sup>In this section (4.2), citations from Georgios Stamatakis are designated to him (GS) by mutual agreement as he was apparently interviewed regarding the oncosimulator. In all other sections, his

As described in the citation, *in silico* cancer modeling following the top-down approach is an iterative process: the more clinical data are supplied to the model, the more accurate it becomes in reflecting reality. However, the top-down approach is very challenging with regard to developing the cancer models. Right from the beginning, the whole range of complexity of cancer has to be accounted for. In order to include multi-scale dynamics in cancer modeling, strategies of how to pass information from a lower-scale level to a higher-scale level and vice versa are required. To solve this problem, each level is characterized by summarizing principles that can be passed to another level of complexity: “That means strategies to summarize what is happening, for example, on the molecular level, and to summarize it in one or two of the very small number of parameters, which can be understood by higher complexity levels. [...] This kind of problems had to be solved in a practical way. And we had to work very hard for this.” (GS)

Instead, a bottom-up approach assembles all parts of a system starting with genes and proteins and brings them into a formal model (Michelson et al. 2006). Therefore, the discovery of each new component needs a reconfiguration of the whole model. Stamatakos (2011, 407) criticizes that this approach focuses rather on microscopic tumor dynamics mechanisms than on multilevel interdependencies and interactions. However, a careful combination of the top-down with the bottom-up approach in the clinical context has its own merits. This has led to the integration of the latter into the latest versions of the oncosimulator.

The oncosimulator envisions to encompass all levels of bio-complexity including the molecular level, the cell level and the supercellular levels. I would say that systems biology in the traditional sense is a very important component of the all-scale approach of *in silico* oncology, or *in silico* medicine in the broader sense. We do need to simulate what is happening on the molecular level, but this is not enough and sometimes the molecular level complexities are so high that you might not even end up with something robust and reproducible. In the real world molecular pathways are very sensitive to crosstalking with other pathways. Therefore, I firmly believe that the molecular level traditionally addressed by systems biology is one of the very important levels to be taken into account in detail. But it is not enough. All other levels should also be taken into account. So you could call such an approach an all-level approach, or a multi-scale approach, or an extended systems biology approach. It is a matter of definition. (GS)

A bottom-up approach focusing exclusively on the molecular level may therefore not deal adequately with concrete and pragmatic questions of importance in the clinical setting. On the other hand, the top-down approach encompassing all bio-complexity levels of the body may be adapted right from the beginning to clinical questions. Of importance are particular questions such as the following. Can the response of the local tumor and the metastases to a given treatment be predicted in size and shape over time? What is the best treatment schedule for a patient regarding drugs, surgery, irradiation and their combination, dosage, time schedule, and duration (Graf and Hoppe 2006)?

The ACGT oncosimulator paves this way by focusing on clinical utility. It primarily aims at supporting the clinician in the process of optimizing patient-specific

---

citations are pseudonomized as are all others.

cancer treatment through conducting experiments *in silico*. Through performing *in silico* experiments, the likely outcomes of several candidate therapeutic schemes are evaluated based on the particular clinical (e.g., symptoms, progress of disease), imaging (e.g., MRI, PET, CT), histopathological (e.g., type of tumor), and molecular (e.g., DNA microarray) data of the individual patient (Stamatakis et al. 2006a).

So let's start from the beginning. The clinical data, the previous treatment history, the imaging data, the body fluid samples, and the biopsy material taken from the patient when available are collected. The extracted multi-scale and inhomogeneous data are pre-processed – some of them through molecular networks, some others by exploiting disciplines such as radiobiology or pharmacology – in order to create the kind of data that the simulation module can understand. At this point, the user, who in the future is expected to be primarily the clinical doctor, describes several candidate schemes, or treatment schedules, or treatment scenarios. They introduce those scenarios into the simulation model. Following the execution of scenarios, the oncosimulator predicts the expected outcome. The outcome is evaluated by the clinician in order to eliminate any eventually not justified extremes or extremely unlikely responses. The user then selects the optimal scheme to be applied to the patient. (GS)

As Stamatakis pointed out in the last quotation, the vision of the oncosimulator is primarily based on its clinical application. Therefore, the cancer models are supposed to be adapted as much as possible to the clinical questions of importance, or in other words, to clinical reality. The predictions aim at supporting clinicians with information on the most effective treatment out of several alternatives, as well as detailed parameters on the optimal composition of a treatment scheme, including the total treatment period, the type of drugs, dose, and interval between treatments (Graf et al. 2009, 147; see Box 4.3).

In consequence, cancer modeling was set up as a top-down approach using all kinds of available clinical data and observations to simulate cancer genesis as a biological phenomenon and its progression under the influence of therapeutic regimes. In this regard, the oncosimulator is not only a clinical tool, but at the same

#### **Box 4.3: Seven Steps How the Oncosimulator Is To Be Used in Patient-Specific Cancer Treatment (Stamatakis 2011, 411f)**

##### **Step 1: Obtain patient's individual multi-scale and inhomogeneous data.**

Data sets to be collected for each patient include: clinical data (age, sex, weight, etc.), possible previous antitumor treatment history, imaging data (e.g., MRI, CT, PET, etc.), histopathological data (detailed identification of the tumor type, grade and stage, histopathology slide images whenever biopsy is allowed and feasible, etc.), and molecular data (DNA array data, selected molecular marker values or statuses, serum markers, etc.).

**Step 2: Preprocess patient's data.** The data collected are pre-processed in order to take an adequate form allowing its introduction into the tumor-and-normal-tissue-response-simulation-module of the oncosimulator. For example,

the imaging data are segmented, interpolated, and eventually fused; subsequently, the anatomic entities of interest are three-dimensionally reconstructed. This reconstruction will form the framework for the integration of the rest of the data and the execution of the simulation. In parallel the molecular data are processed via molecular interaction networks so as to perturb and individualize the average pharmacodynamic or radiobiological cell survival parameters.

**Step 3: Describe one or more candidate therapeutic scheme(s) and/or schedule(s).** The clinician describes a number of candidate therapeutic schemes and/or schedules or no treatment (obviously leading to free, i.e., non-inhibited tumor growth), to be simulated in silico.

**Step 4: Run the simulation.** The computer code of tumor growth and treatment response is massively executed on distributed Grid or Cluster computing resources so that several candidate treatment schemes and/or schedules are simulated for numerous combinations of possible tumor parameter values in parallel. Predictions concerning the toxicological compatibility of each candidate treatment scheme are also produced.

**Step 5: Visualize the predictions.** The expected reaction of the tumor as well as toxicologically relevant side-effect estimates for all scenarios simulated are visualized using several techniques ranging from simple graph plotting to four-dimensional virtual reality rendering.

**Step 6: Evaluate the predictions and decide on the optimal scheme or schedule to be administered to the patient.** The oncosimulator's predictions are carefully evaluated by the clinician by making use of their logic, medical education, and even qualitative experience. If no serious discrepancies are detected, the predictions support the clinicians in taking their final and expectedly optimal decision regarding the actual treatment to be administered to the patient.

**Step 7: Apply the theoretically optimal therapeutic scheme or schedule and further optimize the oncosimulator.** The expectedly optimal therapeutic scheme or schedule is administered to the patient. Subsequently, the predictions regarding the finally adopted and applied scheme or schedule are compared with the actual tumor course and a negative feedback signal is generated and used in order to optimize the oncosimulator.

time a concept of multilevel integrative cancer biology, a complex algorithmic construct, and a biomedical engineering system (Stamatikos 2011, 411). In the next section we look deeper into the model basics to see how the oncosimulator is able to serve as tool, concept, construct, and system at the same time.

#### 4.2.1.1 The Model Basics

At the core of the simulation approach is the idea to explore the natural phenomenon of cancer. In order to describe the biological activity of a discrete tumor spatially, the oncosimulator correlates the response of normal tissue with the response of tumor tissue.

The heart of the system is the tumor and normal tissue response simulation model. This is actually the computer code, a pretty complex simulation code, which gets as input the processed patient data and produces as output the predictions concerning the response of the tumor to concrete candidate therapeutic schemes or schedules (GS).

Based on imaging data, the tumor is simulated as a multidimensional virtual reconstruction including the eventual necrotic region and the surrounding anatomical features before, during, and after treatment (e.g., chemotherapy, radiation). The imaging data provide information on the boundaries of the gross volume of the tumor, the volume itself, and the spatial distribution of the metabolic activity of the tumor (e.g., regions where there is significant provision of oxygen and nutrients through the neovasculature and necrotic regions where there is lack of adequate vascularization and subsequently lack of adequate oxygenation and provision of nutrients).

To simulate how a discrete tumor will spatially spread, the tumor is discretized using a cubic mesh. Each elementary cube of the mesh is called a geometrical cell and is used for the description of the tumor in a statistical way (Stamatakos et al. 2002, 2006a; Dionysiou et al 2004). The geometric mesh covering the tumor region is scanned in certain time intervals (e.g., every 1 h). In each time step, the updated state of a given geometric cell is determined on the basis of a number of algorithms describing the behavior of the cells constituting the tumor. More precisely, each geometrical cell of the mesh belonging to the tumor contains a number of biological cells characterized by the cell phase in which they are found (e.g., stem cells, limited mitotic potential or progenitor cells, differentiated cells, necrotic cells). According to the adapted cytokinetic model (Stamatakos et al. 2006a, 1468), tumor cells usually pass through the following cell phases: G1 (gap 1), S (DNA synthesis), G2 (gap2), and M (mitosis). After mitosis is completed, each of the resulting cells re-enters G1 if the oxygen and nutrient supply is adequate. Otherwise, it enters the necrotic phase which finally leads to cell death. The number of biological cells constituting each phase class is initially determined according to the spatial position of the geometrical cell within the tumor and the metabolic activity in the local area.

It is generally assumed that each geometrical cell of the mesh contains a constant number of biological cells. However, in the case that the actual number of tumor cells contained within a given geometrical cell drops below a given threshold, during the simulation process a procedure starts that attempts to unload the remaining biological cells in the neighboring geometric cells. If the given geometric cell becomes empty it is assumed that the geometric cell is removed from the tumor. Therefore, an appropriate shift of a chain of geometric cells intended to fill in the vacuum leads to tumor shrinkage. This can, for example, happen after irradiation of

a radiation-responsive tumor. On the other hand, if the number of tumor cells in a given geometrical cell exceeds a limit, then additional geometrical cells emerge. By an appropriate shift of a chain of geometric cells towards the boundaries of the tumor, the tumor expands (Kyriazis et al. 2008).

To simulate tumor expansion or shrinkage of a discrete tumor, a number of algorithms (operators) are periodically and sequentially adapted to the anatomic region of interest (Stamatakos et al. 2002, 1771). These algorithms are based on selected parameters<sup>22</sup> influencing tumor growth and response to treatment. They finally steer the simulation as one interviewee explained:

The parameters produce different kinds of data. Therefore, different settings produce different kinds of data. It is like, you know, if you have an oven at home and you have lots of bread that is not baked. You put your oven at twenty degrees, you put in the bread and you see what happens. And then you change the oven to fifty degrees and you put in another bread and you see what happens. And then you put all these breads together again and you see the one that was at twenty degrees is still dough. It's not baked, it hasn't done anything. The one that you put at five hundred degrees is burned. And somewhere in between there is an optimum. This is only one parameter. This is temperature. But there may be other types of parameters that are also going to influence your bread. The humidity or the type of dough that you put in it. How much water did you put in the dough? How much yeast did you put in the dough? Those are different parameters so you can imagine that there are many different kinds of bread. And many different kinds of baking that you can do. And your challenge as a baker would be to find this optimum. Now in the case of oncology, it is not only five parameters, but regarding the *in silico* oncology simulator of Georgios Stamatakos, it was something like forty parameters. So there is a huge space that is spent and that you need to search for the optimum solution. (I3, IT)

Some of the parameter values used for modeling, such as cell-cycle duration or the necrosis rate of differentiated cells, have been based on literature reviews for particular tumor types. Others have been defined based on exploitable medical data and logic. The latter concerns those parameters for which only qualitative data are available. In cases where those different values are available for a given parameter, all values are considered in different instances of the model. Generally only those parameters relevant to the particular type of tumor are used. For instance, in regard to the scenario of pre-operative chemotherapy in nephroblastoma, 30 clinical parameter values are listed covering cell-cycle dynamics to treatment modalities (Graf et al. 2009, 144).

The selected parameters have to be interrelated and ranked based on their effect on the treatment outcome. Some restrictions regarding the interrelation of several parameters (e.g., higher resistance of stem cells to treatment in relation to progenitor cells) are known from the literature. They are further exploited by the simulation process itself.

---

<sup>22</sup> The term parameter is used in mathematics for a quantity or arbitrary constant whose value varies with the circumstances of its application. In a more common sense, parameters are any factor forming one of a set that defines a system and determines (or limits) its performance (Yourdictionary, [www.yourdictionary.com/parameter](http://www.yourdictionary.com/parameter). Accessed July 7, 2014).

The basic idea is that a patient comes in, has a tumor, you use the oncosimulator to predict what the tumor will do when you give treatment, for example, if you radiate it, or if you give it chemotherapy. That has many very different parameters. You can radiate longer or with a higher power. You can radiate daily or semi-daily or maybe weekly. You can give chemo or not. And if you give chemo, there are two different kinds of chemo that you can give. Do you give them both? Do you give them one after another? Do you combine radiation with chemo? Or do it after each other? Or do you first do chemo and then radiation, etc.? So that is a huge parameter space that you build up. Now what we were doing is create visualizations of all these different settings. So you have a number of settings that you keep static. There is one setting that you change a little. And then you can see in the visualization what the shape of the tumor would do. So it could grow. It could stay the same. Or it could shrink. And based on that you could give an indication. Okay, I understand now, if I change this knob in this direction, the tumor will shrink. So okay, this is what I want. This is more optimal. So this is a good change for me. So now I am going to change a different parameter. If you change that parameter, your visualization will change again, but now in two dimensions. So you had this change of one parameter. Then you change a different parameter, so there are two parameters that you change. This is two-dimensional. And then you get different solutions. So you can see what the effect of the second parameter is in combination with the first parameter. If you see this, then you hope to find the pattern. Again, it can basically shrink in a second direction, or it can stay the same in the second direction, or it can increase in the second direction. If you know this you can see what the correlation is of these two parameters based on this visualization and then you include a third parameter, etc. So you combine all these parameters in a visual interface that will show you how the simulation is influenced by changing the setting of one parameter, or two parameters, or maybe three parameters. (I3, IT)

Therefore, further validation, adaptation, and optimization take place following each simulation. The real response of the patient to the treatment is compared with the predicted response and this result is utilized as feedback in order to improve the simulation model. “That means that the more patients that have been addressed by the oncosimulator, the better its predictive potential is expected to become” (GS).

Summing up, the ACGT oncosimulator interprets tumor information according to mathematical measures and models to predict the composition and the shape of the patient-specific tumor and the response to therapeutic regimes over the course of time. Viewing biological processes through the lenses of mathematics means in the first place to quantify living matter that might be affected and determine it in time and space. This has two crucial consequences: this approach relies on as much data as being available, and it relies on parameters that are mathematically applicable.

#### ***4.2.2 Genesis and Development of the Oncosimulator***

The genesis and development of the oncosimulator are deeply embedded in the academic career of Georgios Stamatakos, the teamwork in the In Silico Oncology Group at the Institute of Communication and Computer Systems (ICCS), National Technical University of Athens (NTUA), and interdisciplinary and international collaborations. In the very beginning, after Georgios Stamatakos had passed his master’s degree in bioengineering at the University of Strathclyde (Glasgow, UK) and his PhD in biophysics at the NTUA, the work began with the desire to take off into a new research field and the advice of a supervisor.

Actually, after my PhD thesis, which was on bio-electromagnetics, Professor Nikolaos Uzunoglu suggested to me that I should extend my research interests by doing something regarding radiation therapy. In that period, there was a good collaboration at the Athens Technical University with the Klinikum Offenbach in Germany [...] concerning, for example, the use of electromagnetic fields in order to enhance radiation therapy (hyperthermia). In that way, I started working on radiobiology and radiobiological modeling. But this kind of interaction with the Klinikum Offenbach (Prof. Nikolaos Zamboglou) helped me to get more concrete. And more clinically oriented, let's say. And of course, by trying to utilize a previous expertise in particular I loved the mathematics somehow, I proceeded to the formulation of this concept of the oncosimulator. And then of course, there were a number of students I directed in their diploma thesis or PhD thesis, who all helped to contribute to the implementation of this concept. (GS)

At National Technical University of Athens, the In Silico Oncology Group was set up in 1997 initially working on a number of simulation models regarding tumor response to treatment both *in vitro* and *in vivo*. Again, international interaction and cooperation encouraged Stamatakos and his team to move on with the idea of developing the oncosimulator.

At this point, I would like to particularly mention the very important help we got from Werner Düchting from the University of Siegen in Germany. Actually, Werner worked before us concerning the simulation of tumor growth *in vitro*. His modeling work referred to small tumors. That's before the creation of new blood vessels. And that was pretty inspiring for us to move to the *in vivo* simulation using imaging and multi-scale data as we did. So even before ACGT, there have been quite extensive multinational interactions. I would also call them intercontinental interactions concerning this approach. The contribution of Norbert Graf, professor of pediatric oncology and hematology at the University Clinic of the Saarland, has been of paramount importance. (GS)

In the beginning, the research focused on one specific cancer type. That was glioblastoma multiforme which served as the first paradigm. Afterwards, the In Silico Oncology Group tried to reuse and exploit parts of the algorithms, codes, and the major philosophy of the approach in other cancer types. In addition to glioblastoma, *in silico* simulation has up to now been applied to breast cancer, lung cancer, leukemia, nephroblastoma, and cervix cancer.

From the very beginning, two preconditions have been defined before envisaging the introduction of *in silico* methods into the clinic: first, every prediction of an *in silico* simulation has to be compared with the reality; and second, every *in silico* simulation has to be part of a clinical trial in which the clinical, imaging, biochemical, and genomic data are systematically acquired (Graf et al. 2009, 142). The first simulation using real trial data was based on a clinical trial outcome of the Radiation Therapy Oncology Group, a clinical cooperative group funded by the National Cancer Institute (Philadelphia, USA). Using the clinical study on hyperfractionated radiation therapy and bis-chlorethyl nitrosourea in the treatment of malignant glioma (Werner-Wasik et al. 1996), the In Silico Oncology Group at NTUA simulated the hyperfractionation of two different radiation doses.<sup>23</sup> In regard to shrinkage and regrowth of the tumor, the simulation predicted the real clinical trial outcome in advance (Stamatakos et al. 2006b). The study has revealed that trial participants who received the higher doses had survival superior to the patients receiving the lower

---

<sup>23</sup>One time it was the total dose of 48 gy/day and another time the total dose of 81.6 gy/day.

doses. This was the first breakthrough in the development of the oncosimulator with regard to clinical validation. However, more breakthroughs had to follow. The next step was the development of the first integrated version of the oncosimulator in the course of the ACGT project.

Regarding the oncosimulator, an initial version of the entire integrated system was produced. We started the clinical adaptation and validation process, but this of course will take some years to be completed. Nevertheless, I do believe that we ended up with something pretty concrete. Of course, it is a first version. But still, it is an integrated and complete first version, at least as far as the scientific and technological components are concerned. The clinical aspects, as I mentioned, are much, much more time consuming due to the requirements of the clinical trials, both retrospective and prospective. But as I mentioned, the oncosimulator is being improved and extended within the context of the new projects. (GS)

The initial idea of working on radiobiology has finally resulted in a concrete endeavor of developing a clinically adaptable tool to simulate tumor response to treatment. However, the onward development of the oncosimulator was only possible because Georgios Stamatakos has continuously pursued his vision. Additionally, he has at all times found colleagues who supported his ideas and worked with him together on realizing his vision. In this regard, as we show in the next section, the ACGT project was a very important working environment implementing the oncosimulator into a broader scientific community.

#### 4.2.2.1 Interdisciplinary Challenges

Described as a “really multidisciplinary construct” (GS), the development of the oncosimulator has required expertise from many different domains. First of all, mathematics is cited as being at the core of the whole endeavor. In particular, methods and strategies from discrete mathematics are used to simulate natural phenomena that have a discrete character, for example, the discrete number of tumor cells or the discrete phases of the cell cycle. These discrete entities and quantities in conjunction with their complex interdependences may give rise to predictions of tumor relapse or tumor control over a given time interval (Stamatakos 2011, 409). In addition, strategies of continuous mathematics (e.g., continuous functions, differential equations) are used in order to tackle specific aspects of the models such as pharmacokinetics and cell survival probabilities. Currently more continuous mathematics-based oncosimulators have also been developed.

Since it is a multidisciplinary, scientific, and technological system, it implies that you need, first of all, mathematics. Of course, you need biology. You need expertise in various domains. But mathematics will always light up the heart of these types of systems. And mathematics can be found in any conceivable technological and scientific domain. My view is to try to somehow reuse, extend, and enhance, if possible, already known mathematical methods and, of course, to suggest new ones as well. But at least I personally believe that the effort to somehow extend mathematical methods and tools used in other scientific domains can much accelerate the whole process. (GS)

From a technological point of view, integrating the dynamic and multidimensional visualization of both the medical input and the simulation predictions is

particularly challenging. In addition to virtual reality visualization techniques, further technological components were needed to build a first integrative version of the oncosimulator.

Just to mention that we need components dealing with image processing, internal code parallelization, code acceleration, the execution of the models on several computer architectures including cluster, execution, and nowadays cloud execution, and so on. Grid execution was the one mostly adopted by ACGT. [...] This is a need for the simulator, but there is a need for computer resources, for example, Grid resources. There must be a data management system and, of course, quite a complex interaction of those modules. (GS)

Regarding these scientific and technological tasks and challenges, an interdisciplinary approach was mandatory. The In Silico Oncology Group is therefore composed of scientists with backgrounds in mathematics, informatics, and electrical and computer engineering. In addition, they all need to have an interdisciplinary and visionary mindset.

Of course, there is need for anybody involved in this effort to broaden their horizons. Everybody has to read a lot about scientific fields unfamiliar to them. I would like also to mention the very important contribution of my PhD student, Dimitra Dionysiou, currently a senior researcher in the In Silico Oncology Group, who was actually the first student working on the pre-oncosimulator stage that I directed many years ago. She was very passionate with that idea and, of course, she had to read a lot and do rather unconventional work. (GS)

In addition to the continuous work of the In Silico Oncology Group, interdisciplinary cooperation is often stressed in the interviews. In particular, collaborations within the ACGT project were highlighted in regard to technological solutions of the integrated oncosimulator. One of the interviewees explained how the oncosimulator profited from the architecture of the ACGT infrastructure

So then we moved to more work together with Georgios Stamatakos. He had a problem. His problem was that he had a simulator that would allow you to simulate the effects of a tumor treatment within patients. But the problem was that for every simulation there are a lot of different settings that you can identify that would all make sense, but you would never know which one was the optimal solution. And we had the technological idea that we could use his simulation and sort of try out many different combinations of settings. And use all the computational resources in this Grid infrastructure to do calculations on this simulation in parallel. And then take out the best solution and provide this to an oncologist as a possible treatment for specific patients. That was the idea. So in principle, Georgios Stamatakos, if he had the same question, he would need years of real time to compute an optimal solution. And our idea was to provide an infrastructure that would allow him to do this in, let's say, a couple of hours or maybe even in a couple of minutes by doing all these simulations in parallel and then taking the best solution and provide this, as I said, as a candidate for treatment. (I3, IT)

Even if this cited example illustrates that the oncosimulator has benefited tremendously from the interdisciplinary approach of ACGT, one of the major challenges was in fact to overcome interdisciplinary misunderstanding in the beginning of the project. In the very beginning, for many ACGT partners it was even difficult to understand the idea of modeling cancer.

I still remember the first presentations of Georgios Stamatakos talking about modeling cancer and developing models, etc. Looking back 9, 10 years ago, it wasn't easy for clinicians to understand what he was talking about. You know, can we develop models? Can we model that? I still remember debates trying to use parallels from, let's say, twenty years ago. We couldn't really predict the weather, but we gradually developed models. So we opted for a more systemic approach in trying to develop models that can predict. And today through trial and error we see that our predictions are more accurate or totally accurate. And we had to make each other understand on what we imply on terminologies such as predictive models, develop models of cancer evolution, etc. (I7, IT)

Georgios Stamatakos, kick-off meeting in Nice, he introduces an oncosimulator, he simulates the disease in a computer. Afterward, I went up to him and said, 'what nonsense, what he's doing is nothing but utopia'. That may exist in 100 years, etc. But on the other hand, I found it incredibly interesting, so for me it was something where I thought, okay, let's wait and see. And actually, a pretty good relationship emerged from that, so that we are actually developing this oncosimulator further from the clinical side, and it has become a really close collaboration. (I18, BioMed)<sup>24</sup>

Despite the skepticism in the beginning, the collaboration within ACGT in regard to the development of the oncosimulator was fruitful and lasting. Finally, "I think that practically all members of ACGT were optimistic and we did our best in order to contribute to the shaping and the construction, let's say, of an initial version of this basic science and technology integrative systems biology system" (GS). In particular, Georgios Stamatakos appreciated the optimistic attitude of his ACGT partners, as one of the major problems regarding the development of the oncosimulator was in the beginning of his project the skepticism of other scientific disciplines.

One of the major problems, maybe historically the most important problem, was the reluctance of biologists and clinicians to accept the possibility that such a tool would ever be of clinical use and would ever be translated to clinical practice. And that was not entirely inexplicable in the sense that both biology and medicine at that time were mainly based on empirical knowledge. Of course, there had been a number of biomedical engineering devices, but the idea of bringing together so diverse disciplines and knowledge coming from areas spanning from image processing, let's say, to molecular dynamics or in the spatial scale, let's say, from nanometres to metres. That's in time from nanoseconds to years. That sounded at least in the beginning too futuristic. More a dream than something of any realistic content. Nevertheless, I was not taken back by such a very critical, let's say, approach. (GS)

In this regard, the ACGT project was one of the first working environments where Georgios Stamatakos received approval from the systems-oriented community in oncology. Close and still ongoing interactions and collaboration within the frame of ACGT began during this time. It can be concluded that the first integrative version of the oncosimulator was only possible because of the ACGT environment. In particu-

---

<sup>24</sup>German original: "Georgios Stamatakos, Kick-Off-Meeting in Nice, er stellt seinen Oncosimulator vor, er simuliert die Krankheit im Computer. Da bin ich nachher zu ihm gegangen und habe gesagt, so ein Schwachsinn, was er da macht ist nichts anderes als Utopie. Das gibt es mal in 100 Jahren vielleicht, etc. Aber auf der anderen Seite fand ich das unheimlich interessant, so dass es für mich etwas war, wo ich gedacht habe okay, jetzt gucken wir mal. Und daraus ist eigentlich eine ganz gute Beziehung entstanden, so dass wir tatsächlich diesen Oncosimulator von der klinischen Seite weiterentwickeln, und da ist eine richtig enge Zusammenarbeit draus entstanden." (I18, BioMed)

lar, the collaboration with ACGT partners from the biomedical domain opened the path to work with real clinical data and to start clinical validation on a systematic basis. In retrospect, the Community Research and Development Information Service (CORDIS) of the European Commission highlighted the oncosimulator as one of the EU-funded project success stories.<sup>25</sup>

#### 4.2.2.2 Continuing Research After ACGT

Even if the first integrative version of the oncosimulator was built during the ACGT project and clinical validation has started by using real clinical trial data the oncosimulator was still located in the stage of research after the ACGT project ended in 2010. However, the In Silico Oncology Group was able to continue its work on the oncosimulator in several research projects funded by the 7th EU Framework Program.

In the research project p-medicine, the oncosimulator advanced and expanded in regard to the cancer types (acute lymphoblastic leukemia in addition to nephroblastoma and breast cancer) and the treatment protocols (chemotherapy, targeted therapy, radiotherapy, and combinations). As in the ACGT project, clinical trial data are used in order to optimize and validate the simulation models.<sup>26</sup>

In the research project MyHealthAvatar, the target is not the modeling of cancer or distinct cancer types, respectively, but of personal health in a broader sense. The in silico models of the In Silico Oncology Group are integrated into an ICT infrastructure that aims at collecting, sharing, and offering access to long-term and consistent personal health status data through an integrated in silico environment.<sup>27</sup>

In the research project DR THERAPAT, the digital radiation therapy patient platform is built up to integrate available knowledge on tumor imaging, image analysis and interpretation, radiobiological models, and radiation therapy. The goal is a coherent, reusable, multi-scale digital representation. Radiation therapy was chosen as the application to prove the integration of those concepts because inherently imaging plays a major role in radiation therapy planning and delivery, so the imaging information is available as input for various models, and the delivery process is relatively well understood, making model validation easier compared to, for example, chemotherapy.<sup>28</sup>

In the research project TUMOR that aims at implementing a cancer model repository, the In Silico Oncology Group focuses on multilevel cancer models which address more aspects of the natural phenomenon of cancer.<sup>29</sup>

Finally, the In Silico Oncology Group is the coordinator of the research consortium of CHIC. This research project proposes the development of clinical trial-

---

<sup>25</sup> CORDIS, [http://cordis.europa.eu/result/brief/rcn/6061\\_en.html](http://cordis.europa.eu/result/brief/rcn/6061_en.html). Accessed June 1, 2014.

<sup>26</sup> p-medicine, <http://www.p-medicine.eu/>. Accessed June 1, 2014.

<sup>27</sup> MyHealthAvatar, <http://www.myhealthavatar.eu/>. Accessed June 1, 2014.

<sup>28</sup> DR THERAPAT, <http://drtherapat.eu/>. Accessed June 2, 2014.

<sup>29</sup> TUMOR, <http://tumor-project.eu/>. Accessed June 2, 2014.

driven tools, services, and infrastructures that will support the creation of multi-scale, integrative cancer models. One important focus is on the standardization of model description and model fusion. The creation of such elaborate and integrated models is expected to sharply accelerate the clinical translation of multi-scale cancer models and oncosimulators following their prospective clinical validation.<sup>30</sup>

Because of the collaboration on a range of research projects funded by the 7th EU Framework Program, the In Silico Oncology Group was able to include further cancer types (e.g., leukemia, lung cancer, prostate cancer) into the oncosimulation as well as further treatment protocols. In addition, clinical validation of the oncosimulators advanced according to the increasing access to clinical trials and real patient data. From a theoretical point of view, multiscale cancer modeling progressed towards more and more integrative models. In particular the last listed research project CHIC indicates this next step in the development of in silico oncology. The so-called hyper-models are defined as choreographies of component models, each one describing a biological process at a characteristic spatiotemporal scale. The component models are related to hyper-models defining the relations across scales and integrative models can become component models for other integrative models (Stamatakos et al. 2013).

In CHIC, the next steps of in silico oncology are already targeted: the development of an infrastructure that will support accessibility and reusability of mathematical and computational hyper-models. The standardization of cancer model and data annotation allowing multiscale hyper-modeling is one of the preconditions to be fostered in the future. In addition, the secure access to already existing data, models, and analysis tools is estimated as a necessary requirement in the development of in silico oncology. Accordingly, the set-up of extensive, in silico oriented repositories (e.g., hyper-models, hyper-model driven clinical data, distributed metadata, in silico trials) are demanded to keep track of the development of simulating cancer in silico.

### 4.2.3 Concluding Remarks

The description of the ACGT oncosimulator has shown that the development of an innovative technology is a story deeply connected with very different incidences. The most important step of the oncosimulator's storyline is, of course, its beginning: the initial idea to support clinicians with predictions on the most effective treatment out of several alternatives. This ultimate research objective has come up very early in the emerging field of in silico oncology. Just as important as the vision itself is the initiator, Georgios Stamatakos, who has consequently developed the first idea and expanded it to a vision of a biomedical technology that entails a comprehensive concept of multilevel integrative cancer biology, a complex algorithmic construct, and a biomedical engineering system. Furthermore, Stamatakos has at all stages of

---

<sup>30</sup>CHIC, <http://www.chic-vph.eu/>. Accessed June 2, 2014.

research found colleagues who supported his ideas and accompanied him on the long path realizing his vision.

Fundamental to the ongoing story is also that Stamatakos and his team pursued the initial version of the oncosimulator and, from the beginning, have subordinated any decision made in the course of research to this original vision. Even though this basic philosophy has sometimes led to take very difficult tracks, Stamatakos and his team pursued their ideas. For instance, the vision of the oncosimulator is primarily based on its clinical application. Therefore, the cancer models are supposed to be adapted as much as possible to clinical reality. As a consequence of this, cancer modeling was set up as a primarily top-down approach using all kinds of available clinical data and observations to simulate cancer genesis as a biological phenomenon and its progression under the influence of therapeutic regimes. At this stage of research, Stamatakos and his team have already reached at one of the core challenges of systems biology: the multilevel integration of biological processes. Again, Stamatakos chose a pragmatic approach to overcome this problem. To move on in the development of the oncosimulator, each level was characterized by summarizing principles as a set of parameters that can be passed back and forth between different levels of complexity.

Another far-reaching decision made at the start was to give priority to strategies and methods of particular mathematics. By viewing biological processes through the lenses of particular mathematics, all natural phenomena that might be affected were quantified and as much data being available are collected and used in constructing the models. The formalization of those mathematic models is realized by ICT. Even though real clinical data play such a prominent role in oncosimulation, the mathematization and formalization of biological processes at the same time limits their digital reconstruction. In the models, those phenomena are primarily being considered which have a discrete character, for example, the discrete number of tumor cells. In addition, the mathematization and formalization limit the application of parameters that are basically steering simulation and prediction. In the simulations, only those parameters are being considered that are mathematically and digitally applicable. To sum, the analysis of the oncosimulator's underlying conceptualization has shown that the consistent application of mathematic modeling formalized by ICT has far-reaching consequences on doing research: those concepts (e.g., discrete character, applicable parameter) shape the research process from scratch and restrict it at the same time.

The use of mathematical methods and tools in other scientific domains and its ascribed supremacy is possibly the most important reason why interdisciplinary problems have occurred in the course of the oncosimulator's development. However, Stamatakos was able to overcome the reluctance of biologists and clinicians, described as historically the most disturbing factor in the storyline, by convincing them step by step. It started with individual scientists in the ACGT project and is still ongoing in continuing research after ACGT. EU funding seems to be a good environment to meet and collaborate with interdisciplinary-minded scientists and to build up international communities on the edge of emerging research fields such as *in silico* oncology.

### 4.3 Coordinating Systems-Oriented Research in a Technological Environment

The quantity of data involved has generated the idea that data-intensive science is a whole new way of doing research (e.g., Mayer-Schonberger and Cuckier 2013; Kitchin 2014). Although the production of such large data stocks has already existed in some domains for some time (e.g., weather prediction, financial markets), biological research has transformed into data-intensive science in the context of Omics and systems-oriented research since the early 2000s (Leonelli 2014, 3). This coincided with the time when technologies for the high-throughput production of genomic data (e.g., DNA sequencing, microarrays) started to become widely used.

The large data stocks have made it necessary to reorganize the storage and management of data. Expectations regarding the use of ICT in systems research to manage large data repositories are currently high and ICT infrastructures are already under way towards realization. The following section focuses on the discrete results or outcomes of the ACGT research project to retrace the current status of ICT in systems-oriented research and to assess the potentials of such an approach. Based on empirical data, we tried to find out what the ACGT members thought to be necessary to maintain an ICT infrastructure and to keep it running and how the scientists evaluate its productivity. Questions related to such issues were included in the questionnaire we used in the interviews described in Sect. 4.1.1. We asked *inter alia* for the concrete results of the ACGT research project and the reasons why the goal of designing an ICT infrastructure and implementing it into the emerging systems-oriented research community in oncology was not fully reached when the ACGT project was concluded. Many interviewees agreed that the ACGT project was not able to accomplish an ICT infrastructure that can be used in clinical practice because it requires much more effort in terms of financial support and time to be invested than was available in a four-year research project. Therefore, they broadly discussed what lessons they learned regarding the use of ICT for systems-oriented research. The question is: what kind of function will ICT infrastructures have, or, in the eyes of the interviewed ACGT consortium members, are they supposed to have in systems medical research?

#### 4.3.1 *The ACGT Project and Its Results*

In the proposal of the ACGT research project, its objective was clearly defined: the ACGT consortium aimed at designing and developing an integrated ICT infrastructure that offers tools and techniques for the mining of data from data repositories and the extraction of knowledge from knowledge discovery services (see Sect. 4.1.1). Hence, the project's results can be directly compared to and evaluated by the objective described in the proposal. However, the following section shows that the interviewees consider not only technological innovation, but also indirect outcomes such as gaining experience in the research process as valuable results of the ACGT project.

### 4.3.1.1 Technological Innovation

As the research guiding objective was the creation of an integrated ICT infrastructure, it is certainly not surprising that the whole endeavor was described as a technological innovation starting from the outset.

“In the beginning, there was nothing,” explained an IT-expert in the interview (I3); “so we had an objective and the objective was to create an infrastructure that would allow you to do scientific research over a distributed platform. A platform that would consist of many different institutions that all had their own computational resources that would allow you to do research that was not possible before. But in the beginning we did not have any infrastructure so this Grid infrastructure needs to be created.” The computer scientists within the ACGT project started by investigating what type of software was available and what kind of conceptual systems were already built and how existing databases worked. After the basic decisions were made of how to create the ICT platform, the assigned ACGT partners developed different technological components and tools such as the data access services, the clinical trial management system, or the workflow editor (see Sect. 4.1.1). These components were composed as parts of an integrated system. However, many problems occurred when the components designed by different ACGT partners were to be assembled into an integrated architecture.

The general idea how the components were supposed to interact, that existed already. But whenever we sat down together, when we programmed something together, linked up a few things from various partners, then there was always some kind of problem. And then it sometimes took weeks to find out what the problem was. That was also a reason why ACGT wasn't so successful, because this Grid technology is very, very complex. That means, in the following project we're not taking that kind of approach any more. Instead, we're trying to keep things simple. Because it really may be that in the end, the problem is ... if a workflow doesn't run properly because, say, the computer on Crete, that computer's clock is a millisecond ahead of the clock we have here. And then some security alarm went off because it thought that data from the future are coming in—that can't be, so it aborted the process. But you've got to figure that out, and it isn't easy. That can take days and weeks until you've figured out somehow, going through the entire system why one part somewhere seems to think that something isn't working anymore. (I12, IT)<sup>31</sup>

In the course of the ACGT project, the coordination and assembly of the components was continuously presented as an end-to-end demonstration at meetings in front

---

<sup>31</sup> German original: “Die allgemeine Idee, wie die Komponenten zusammenspielen sollen, gab es halt. Aber es war immer so, wenn wir uns zusammengesetzt haben, irgendwas zusammen programmiert haben, ein paar Sachen von verschiedenen Partnern verknüpft haben, dann gab es immer irgendwo ein Problem. Und dann brauchte man teilweise Wochen, um herauszufinden, woran es lag. Das war auch ein Grund, warum ACGT nicht so erfolgreich war, dass diese Grid-Technologie sehr, sehr komplex ist. Das heißt, im Nachfolgeprojekt haben wir so was auch nicht mehr, sondern versuchen das einfacher zu machen. Weil es wirklich sein kann, dass im Endeffekt das Problem daran ... wenn ein Workflow nicht durchläuft, weil die Uhr, die irgendwie der Rechner auf Kreta hat, irgendwie eine Millisekunde vor der Uhr läuft, die wir hier haben. Und dann knallte irgendwas mit der Security. Weil der meint, da kommen irgendwie Daten aus der Zukunft—kann nicht sein, und bricht ab. Und da muss man halt erst mal drauf kommen. Das kann halt Tage und Wochen dauern, bis man dann irgendwie durch das gesamte System herausgefunden hat, warum irgendeine Stelle meint, dass jetzt irgendwas nicht mehr funktioniert.” (I12, IT)

of reviewers assigned by the EU commission.<sup>32</sup> These demonstrations were adapted to a scenario-based development process in which a number of scenarios were created. Essentially, they can be described as a sequence of activities conducted by a clinician who is willing to use the ACGT platform in his or her clinical trial. The sequence followed the established procedures of data handling in a clinical trial, that is, access to heterogeneous data, use of various tools for data analysis, and invocation of appropriate tools for visualizing and interpreting results (e.g., ACGT 2009).

The clinician as the final end-user of the ACGT infrastructure was in focus of the scenarios. However, it was often stressed in the interviews that the ACGT project was a research and development project (R&D project). After only four years of research, the developed infrastructure was not ready for regular use in clinical practice and many of the interviewed members of the ACGT consortium did not initially anticipate that by the end a sustainable infrastructure would exist that could be of use to the oncological community. In their view, the ultimate objective of the ACGT project was just to prove the concept. They wanted to show that developing an ICT infrastructure for clinical systems research in cancer is possible: “ACGT was a kind of a proof of concept. As is the case I think of the most of EU projects. You are trying to build something to show that it is possible and of course, you are trying to build up on it in future projects. And try to reuse it. But it’s not building a production level system.” (I5, IT) According to the quotation, clinical application (“the production level,” previous citation) was not the scope of the ACGT project but the proof of concept which means in the first place to develop an infrastructural prototype. This is what the ACGT consortium achieved: the first integral version of an ICT infrastructure was presented as an end-to-end demonstration at the final review meeting held in Heraklion (Crete) in September 2010.

Concerning the technological outcomes, it can be said that not the infrastructure, but the individual tools such as the clinical trial management system called OpTiMA or the oncosimulator were the most concrete technological achievements of the ACGT project. Many of those components hold the potential for further use in follow-up projects. For example, the security tool named the Custodix Anonimisation Tool that supports anonymization and pseudonymization of different types of data, designed by the software development company CUSTODIX, is already reused and extended in follow-up projects.<sup>33</sup> The integrated ACGT infrastructure itself broke down several months after the research project had ended. The reason was a very practical one: the technical partners switched off the server capacities for the ACGT infrastructure one after another, and the ACGT computing network that was built all over Europe broke down.

---

<sup>32</sup>End-to-end demonstration basically means that the assembly of components into a system is demonstrated by creating workflows using the system from end to end.

<sup>33</sup>The successor CATS is a versatile service platform for de-identification and pseudonymization which can easily be integrated into high-volume data workflows and is, for example, applied in the ACGT follow-up research projects p-medicine ([www.p-medicine.eu](http://www.p-medicine.eu)) and INTEGRATE ([www.fp7-integrate.eu](http://www.fp7-integrate.eu)). Both websites accessed September 15, 2014.

### 4.3.1.2 Experience

The ACGT platform did not persist; however, the project instigated research on ICT infrastructures in the biomedical domain by many former members of the ACGT consortium. In reference to the broader research field, the interviewees highlighted in particular the experience and the knowledge they gained in the research process as a valuable outcome. The ACGT project was “a very good basis for the things that we are doing now” (I2, IT) said an interviewee who is currently working in one of the follow-up EU projects. The ACGT project therefore seemed to be a starting point of promising research that was worth being pursued in future work.

So in a general point of view, I suppose that we gained a lot of understanding on how difficult it can be to create an infrastructure that is very technological, at a very bleeding edge, advanced, and apply that to a setting that has no clear understanding of the computer science ideas behind it. And so, there were a lot of difficulties that we had to overcome. But, you know, during the course of the project, we also gained a lot of understanding on how we could cope with those difficulties and how we could sort of fix the underlying challenges. That is one. And the other thing from our own personal perspective is that we created something new. We created an architecture that we still use, not in the same types of projects that we did with ACGT, but we are now applying the same type of research to other projects. It allows us to continue the research that we have done and extend on it. So that is good from our perspective. I suppose, but it is guessing, I suppose that from the clinical point of view, there is a better understanding of how computer science can help the research in a clinical setting. Especially also on subjects that have to do with, let's say, the genetic backgrounds and everything that has to do with proteomics, the Omic types of research. We are not a part of p-medicine, once again, but I suppose that the people in p-medicine have a clear understanding of how they could continue with the work, the results that were produced in ACGT. And how you can build upon that and get your own science further and better. (I3, IT)

Assessing ACGT's impact on future research, the cited interviewee underlined that the scientists were gaining a deeper understanding of the theoretical challenges and practical obstacles of developing an ICT infrastructure. This means that the awareness of the problems was created by practical experience in the first place. This approach was expressed by others as well. For instance, biomedical experts pointed out that they learned more about the possibilities and limits of ICT. Some of them considered for the first time the ethical–legal requirements that are indispensable when designing clinico-genomic trials and having a continuous access to data-sharing platforms. Computer scientists, on their side, got an inside view into daily clinical workflows, the amount of information that can be generated from genomic data, and how sensitive these data are in legal and ethical terms.

This experience was described as being important for future research in follow-up projects: “A lot of things were used in p-medicine. Yes. And mostly the experience was used, which is, in terms of time, huge. The biggest thing a lot of times is the experience of what the problems were rather than the actual building of the tool, because the actual building of the tool doesn't take that long if you know exactly what you need” (I9, BioMed).

Looking at ACGT as seen and assessed by the interviewees, it can be concluded that the concrete results of the project were primarily technologically defined.

However, only individual components such as the clinical trial management system, the security tool, or the workflow editor were positively assessed as having the potential for further use, but not the integrated architecture of the ICT infrastructure. Yet, personal outcomes were especially discussed to be as important as the technical ones. The ACGT members appreciated that they were gaining experience in the emerging field of systems-oriented research in oncology. They mentioned in particular that they deeply explored technological and theoretical concepts (e.g., Grid computing, ontologies) and gained practical experience in interdisciplinary work (e.g., clinical workflows, data protection standards) or got opportunities for interdisciplinary networking. These more indirect outcomes were regarded as having a crucial impact on future research in systems-oriented research.

### **4.3.2 *The Sustainability of ICT Infrastructures***

Several months after the ACGT project was finished, the integral ACGT infrastructure was shut down. Most of the former ACGT members were not willing to provide server capacities for an indefinite time for an R&D project that was already terminated. In addition, the services were often needed and reused in follow-up projects. This is not unusual in research projects, as one IT expert explained: “The fact that it kept on running before it was finally turned off, that it was in sleep mode, in Halbernet mode, that’s what’s unusual. The individual institutes, there’s no way they can achieve that simply because they’re research institutes” (I14).<sup>34</sup> Accordingly, the interviewees collectively agreed that one of the basic problems why the ACGT platform did not succeed to be used in the clinic was the lack of sustainability with regard to server capacities after the completion of ACGT. Claims for sustainability were often expressed in the interviews. We therefore aim at exploring how sustainability affects the potential of ICT in systems-oriented research. The analysis shows that sustainability is defined not only for server capacities, but for different objects and different contexts. In the following sections, we take a closer look at those objects and contexts that, in the eyes of the interviewed scientists, need to be sustainable to keep ICT infrastructures running and implement ICT into systems research.

#### **4.3.2.1 Technological Sustainability**

To keep ICT infrastructures running and finally to implement ICT into systems-oriented research, the ACGT interviewees regarded sustainability defined in technological terms as *sine qua non*. In this context, the technical design of the prototype itself was criticized by some interviewees. They discussed why the ACGT prototype

---

<sup>34</sup> German original: “Dass es, bevor es endgültig abgeschaltet wurde, so lange noch im Schlafmodus, im Halbernet-Modus, weiter gelaufen ist, ist eher das Ungewöhnliche. Die einzelnen Institute, die können das gar nicht leisten, weil es halt Forschungsinstitute sind.” (I14)]

was in their view not the best solution to build up a user platform for the systems-oriented research community in oncology. One reason given was the software used that was still in its infancy.

A lot of the technology that was used within ACGT was not mature enough. But those pieces were not generated within ACGT. So that is unfortunate, because you are basically building on something that is not mature yet. But you are trying to assess if even this immature technology can be applied to your context. So that is the research project and I think that the results that came out of ACGT were very enlightening. Because a lot of progress has been made on understanding what did not work and what did work. So you can use that in the next project. You now use the things that worked and try different things for the things that did not work. But you are always faced with pieces of the technology that are outside your power. Right so, in ACGT, there was a lot of Grid software that was used, that was developed in other projects like Aggie or Cern, in other European research projects that were primarily focused on creating this Grid infrastructure. So there is nothing you can do about this. (I3, IT)

As the quotation indicates, in particular, the software used came under criticism for its immaturity and complexity. The Grid software is complex as it ties the technical components of the infrastructure closely together.<sup>35</sup> At the same time, the infrastructure was distributed. This means that each technical partner of the ACGT project was requested to offer one or more server(s) that would then be connected with the servers of the other ACGT partners. Hence, the system as a whole and not only the parts of it had to be maintained for its sustainability. As one interviewee put it, “a key question that was set in the beginning of the ACGT was the following: is there value in such a setting of employing a Grid infrastructure? Which I think it is one of those cases where you spend a lot of effort in trying to find the answer and the answer at the end of the day is, it is probably not” (I7, IT). At the same time, Grid technology is more and more being replaced by Cloud technology as computers are becoming faster and cluster computing is no longer necessary. Of course, computer scientists are familiar with such technological developments in which one approach is replaced by another. Looking into their daily work, IT specialists seek cutting-edge technologies such as Grid or Cloud computing in order to use these technologies for their designated application, such as designing an ICT platform for systems-oriented research in cancer.

Seen from today’s perspective, the ACGT project was not only working with immature technology, but the Grid computing approach itself was questioned and even soon outdated. Drawing on this argument, one of the interviewees explained that the concepts are still the same although the underlying software might change.

[W]e are not talking about Grid systems anymore, but we are talking about Cloud systems. And it is a subtle change in approach, but the technology questions are still the same. [...] It mostly boils from the awareness that the Grid technology that we created was far too immature. So they ripped out portions of it and they took different portions and integrated that and now it is called the Cloud systems instead of the Grid system. But the concepts are

---

<sup>35</sup> Grid computing is based on the collection of computer resources from multiple locations to work on a common goal or project.

still the same. And from my understanding, once again I am not part of p-medicine, but from my understanding they are now applying Cloud technology in p-medicine. That is a logical approach. It makes sense. (I3, IT)

However, the interviewees coincidentally said that the ACGT platform was not implemented in systems-oriented research in oncology because it was still a prototype. Many interviewees expressed that the final, but crucial step was not reached during the project: the step from experiment into practice. Hence, the prototype needs to be converted into a production system that can be regularly used by clinicians who are not familiar with high-performance computing. To be ready for customers, engineering of the research software is necessary. This means that the software has to be tested and consolidated, documented, and, finally, certified.

Research software lacks things that a real application has, like error tolerance, user interfaces, menus or manuals, well, completely normal trivial things that are totally uninteresting for a research project. For example, you show the prototype in a review, that's a proof of concept. You say, this is what we have in mind, this is how it's supposed to work. This is how it would work, fundamentally speaking. That works now. But to be able to sell it, practically as a system, well, quite a lot is still missing, namely software engineering. That means that you have a test department of your own. That means that there are people on staff who really test things from morning to night, checking the whole thing for bugs. There are people for documentation. You don't have them, either, in a research project. The deliverables where you could say, well, a lot of text was produced about the tools, they're for real end users who weren't involved in the project, hardly comprehensible or useless. Well, those are things that are really missing and that take a whole lot of time. And in the world of research, that often isn't so clear. (I14, IT)<sup>36</sup>

From the quotation it follows that scientists coming from university are often not familiar with the requirements of a tool expected to be ready for application on the market. Another interviewee outlined that he realized in the course of the project that it was impossible to establish the ACGT infrastructure for clinical use because of the lack of financial support and time to be invested into software engineering and marketing to achieve marketability.

It's a vision... and I was ambitious together with a number of other people. Not everybody, but a number of other people. But at the same time, you need to be aware of what the reality is and what life is. And having gone through close interaction of what had caBIG achieved in the United States, I had discussions and I had meetings with the director of caBIG,

---

<sup>36</sup>German original: "Forschungssoftware fehlt für eine richtige Anwendung solche Sachen wie Fehlertolerabilität, Benutzerschnittstellen, Menüs oder Handbücher, also ganz normale triviale Sachen, die für ein Forschungsprojekt völlig uninteressant sind. Den Prototyp zeigt man zum Beispiel in einem Review, das ist ein Proof of Concept. Man sagt halt, wir stellen uns das so vor, so sollte das laufen. So würde das grundsätzlich gehen. Das funktioniert jetzt. Aber um so etwas quasi als System verkaufen zu können, da fehlt noch wirklich relativ viel, nämlich Software Engineering. Sprich, dass man ein eigenes Test-Department hat. Sprich, dass halt Leute da sind, die wirklich von morgens bis abends nur testen, das Ganze nach Bucks durchschauen. Es gibt Leute, die dokumentieren. So was fehlt auch in einem Forschungsprojekt. Die Deliverables, wo man sagen könnte, es wurde ja viel Text produziert über die Tools, die sind für richtige Endnutzer, die nicht in dem Projekt drin waren, kaum zu verstehen oder unbrauchbar. Also das sind Sachen, die wirklich fehlen und die sehr viel Zeit brauchen. Und das ist oftmals in der Wissenschaft nicht so klar." (I14, IT)

Professor Buetow, etc. They had a structure. They had offices. They had a marketing director. They had scientific directors. They were functioning as a kind of a company whose task was to develop, to further develop, to open new directions for additional work. But also to make sure that there is support for the community to publicize, to market, etc. And they had the 10 million minimum per year to support their functioning, etc. When you compare that to a European R&D project, although the ambition and the vision was there and I think were supported very heavily, very nicely through our reviewers [...], we realized that it cannot happen. It is very rare that you see a European R&D project, because it is an R&D and not a development project that you end up with a fully functioning infrastructure and the reason for that is that... there are three reasons. Because up to the very end you are exploring scientific and technical issues so you are doing research at various levels. The second is the fact that in European R&D projects, you develop proof of concepts and not production quality systems, the third is that very often you see research groups, once they have reached the proof of concept prototype and published, they lose interest in making it in production quality and production ready system. (I7, IT)

The quotation again referred to ACGT as an R&D project entering new research territories at various scientific and technological levels but not the market. However, several private companies that are usually familiar with the adaptation of products to the market were integrated into the project. Would it have been possible that these companies focus on marketability or how do they define their role in R&D projects such as ACGT? One of the interviewees explained that companies pursue their own interests why or while they seek to be involved in academic research. Essentially, they participate in order to understand trends in future research and to be involved in innovative developments. "First of all, for us it is a kind of an early warning system. We get to listen to academics and what they think is the next big thing although often we find that we tell academia where things are going. It is good, because these things definitely give you a very good platform to project yourself and be seen as an avant-garde company so that you are involved in new things, state of the art things" (I4, IT).

This interviewee assigned the potential of trend-setting innovations to academic research, although companies seem to play a role in the second attempt. Another interviewee of an internationally oriented enterprise pointed directly to the commercial sector and how this would influence his own work.

The alternative would be that you let industry make a decision. So you go, for example, to Microsoft, and tell them this is my problem and please advise me. Then Microsoft will create a Windows Cloud or Windows Grid or something like that. But it will not give you the opportunity of influencing the decisions that are going to be made there. So you are basically forced to swallow the decisions that Microsoft would have made if they decide to build something like this. Therefore, you have to conform to what they did. Whereas in research projects, there is still a possibility of saying to people who have developed the technology, 'the decision that you made there is maybe appropriate for your line of research, but it is not appropriate for my line of research so please can we talk. Can we figure out a way of trying to solve this?' (I3, IT)

From this it follows that research projects, in particular R&D projects, open up space for new trends and approaches in research. In fact, ACGT was one of the first projects exploring how Grid technology could be applied for doing oncological research. Even if the integral ACGT platform was not mature and sustainable enough to reach clinical use, the interviewees collectively agreed that the results of

the ACGT research were valuable and necessary and provided a basis for understanding key issues such as data integration or sustainability of ICT infrastructures. At the same time, the innovative processes taking place seem to be open or democratic enough to allow different stakeholders (e.g., academia, industry) to develop and influence landmark decisions for future research. However, neither academia nor industry seems to be willing or able to be responsible for market introduction within an R&D project. This last aspect is extremely important for the translation of systems biology knowledge and tools to applied research such as systems-oriented research in oncology.

#### 4.3.2.2 Financial Sustainability

Another element of sustainability discussed in the interviews of how to make an ICT infrastructure sustainable to keep it running was the funding. A central server as a sustainable facility where the software can be hosted to allow research to continue after an R&D project had ended was one suggestion mentioned in the interviews. However, to host an ICT infrastructure requires continuous financial support and manpower for maintenance tasks. In this context, some of the interviewees stressed that to date researchers (and funders) usually have the mindset that you don't have to pay for Internet use. To solve this problem, a new path of institutionalization is currently being developed in the ACGT follow-up research project p-medicine.

A structure is established that is going to be dynamic and that will adapt to new circumstances in the future, too. But it's supposed to be a structure where I can continue to do this research. The important thing is to assemble data, to evaluate them, and to put them in a system that can continue to exist independently of EU funding. That means, we're currently trying to develop a business plan where we, for example ... a very simple example. If you take OpTiMA, it's structured like a modular system. If I take this Trial Outline Builder, then you can set it up so that I can collect data without having this Trial Outline Builder. You can get a basic module in OpTiMA for free. And then, if somebody wants to have this Trial Outline Builder, then they can buy it via licensing fees etc. If the modules that I can attach to it are so attractive that someone says, that's what I need, then they'd buy it via licensing fees. If I use a data management system at the hospital, I have to pay for that, too. Well, we're trying to establish long-term funding with this kind of ideas for a business plan. (I18, BioMed)<sup>37</sup>

---

<sup>37</sup> German original: "Es wird eine Struktur aufgebaut, die dynamisch sein wird und die sich auch in der Zukunft wieder an neue Gegebenheiten anpassen wird. Aber es soll eine Struktur sein, wo ich diese Forschung weiter betreiben kann. Das Wesentliche ist eben Daten zusammenzuschweißen, die auszuwerten und in ein System einzubringen, was auch weiter bestehen kann unabhängig von einer EU-Förderung. Das heißt, wir versuchen im Moment einen Businessplan zu entwickeln, wo wir dann zum Beispiel ... ein ganz einfaches Beispiel. Wenn man OpTiMA hat, das ist ja aufgebaut wie ein modulares System. Wenn ich diesen Trial Outline Builder hole, dann kann man das so machen, dass ich Daten sammeln kann, ohne dass ich diesen Trial Outline Builder habe. So ein Basismodul in OpTiMA gibt es dann umsonst. Und wenn jemand dann aber diesen Trial Outline Builder haben will, dann kann er sich den dazu kaufen über Lizenzgebühren etc. Wenn diese Module, die ich dann da dran hängen kann, so attraktiv sind, dass jemand sagt, das brauche ich, würde er es dazu kaufen über Lizenzgebühren. Wenn ich heute irgendein Data Managementsystem benutze in der Klinik, muss ich auch dafür zahlen. Ja, und über solche Vorstellungen eines Businessplans versuchen wir eine langfristige Finanzierung zu etablieren." (I18, BioMed)

The license fee provides the possibility to afford staff for data management, including data curation, and for advancing the ontology implemented into the ICT infrastructure. The business plan mentioned by the interviewee refers to the institution named the Study, Trial and Research Center (STaRC) which is currently on its way to becoming an innovative center to host and provide a service-oriented clinical research infrastructure based at the University Clinic of the Saarland in Germany.<sup>38</sup> The researchers will have the opportunity to run clinico-genomic trials and do systems-oriented research on the STaRC platform by paying for its use. However, the researchers have to take up the offer by actually using this ICT infrastructure. As clinicians are accustomed to paying for the use of data management systems in the clinic, the license fee, as outlined in the above quotation, will rather be contextualized in data management systems than in Web-based services.

Most of the interviewees are positive about future research because the results achieved in the ACGT project will be taken to the next level of realization in the follow-up research projects. It thus appears logically that the institutionalization of STaRC breaks new ground in different directions, not only in research but also in academic mindsets to create financial sustainability.

#### 4.3.2.3 Social Sustainability

To break new ground in research and to put innovation into practice always has a social dimension. It requires scientists who change or widen their mode of thought and of doing research. At least the latter aspect is deeply embedded in social interactions as one interviewee of the biomedical domain outlined. Before participating in the ACGT project, the interviewee did mostly clinical research and only worked together with clinicians. Today, he is working with researchers of different disciplines to translate systems-oriented approaches into clinical practice. In the following citation, he is convinced that only cohesive interdisciplinary teams will be able to improve survival rates and progress in health care.

I have the feeling that you can really make things happen here if you can get everyone with a say in the matter to the table. And it really isn't just the medics who can treat patients in the end. In the future, they'll need IT people. They'll need systems biology. They'll need ethicists and lawyers. They'll need basic research. They'll need the bioinformatics people. And in the future, you'll only be able to help a patient if you have a cohesive team like that. Take pediatric oncology: in the last 30, 40 years, we've achieved a really steep increase in survival rates. We achieved that by working together, doing prospective clinical studies, and gaining new knowledge to improve therapies. And that worked, for purely, ... well, clinical considerations. Then, molecular biology was added to characterize patients better. And nonetheless, we're stuck when it comes to certain groups that we can't get healthy. That means that the steep increase in improving survival rates has been turning into a plateau for

---

<sup>38</sup> STaRC was founded by Norbert Graf, the director of the Clinic for Pediatric Oncology and Hematology at the University Clinic of the Saarland and the former quality manager in the ACGT project. Start-up financing is provided by the federal state government of the Saarland and the European Union in the course of the research project p-medicine in the 7th Framework Program. See STaRC, <http://eu-starc.eu>. Accessed October 3, 2014.

about the last 10 years. Suddenly, we can't improve a certain survival rate and we don't know why one patient is relapsing and another one isn't, because our current knowledge is the same for both. That means we're lacking information. And we have to get that information from these approaches. That's the only way, namely by putting together really all the data about the patients that you have. By developing disease models with the systems biology approach and then combining them and finding out individually for each patient what the best treatment is. And that's the reason why that's a very important development for clinicians. (I18, BioMed)<sup>39</sup>

Of course, for the interviewee cited it was easy to meet and collaborate across disciplines as research projects targeting the development of an ICT infrastructure are interdisciplinary positioned. Many of the interviewed consortium members depicted the ACGT project as a starting point for continuous interdisciplinary cooperation. "So we actually formed some kind of team" (18, IT), said an interviewee who collaborated with nearly all of his former ACGT colleagues afterwards. Someone else stressed that not only collaboration between people but also between the institutions are of vital importance and require continuation. Here, the impact of ICT becomes evident as continuous relations between research centers are not based on personal relations any more.

The cooperation was crucial I think, not only between the persons, but the centers. The information was written down, but it would challenge a similar group of people to go through all the information and learn the same lessons. It is true that you write down the information, but to really read it and use it all would probably take one year on its own, all that amount of information. So in terms of time it was very good that the same people worked on it, because they had both the experience and the access to the same tools. I mean the biggest thing with ACGT is that within four years it created a link between centers and people that never spoke with each other before. Some of them did, but a lot of them didn't. (19, BioMed)

---

<sup>39</sup>German original: "Ich habe das Gefühl, dass man hier tatsächlich etwas bewegen kann, wenn man alle die Leute, die was zu sagen haben, an einen Tisch bringt. Und es sind eben nicht nur die Mediziner, die am Ende Patienten behandeln können. Die brauchen die IT in Zukunft. Die brauchen die Systembiologie. Die brauchen Ethiker und Juristen. Die brauchen Basic Research. Sie brauchen die Bioinformatiker. Und nur, wenn man so ein geschlossenes Team hat, wird man tatsächlich in Zukunft einem Patienten helfen können. Wenn ich mir die Kinderonkologie angucke, dann haben wir in den letzten 30, 40 Jahren einen ganz steilen Anstieg von Überlebensraten bekommen. Das haben wir dadurch bekommen, dass wir zusammengearbeitet haben, prospektiv klinische Studien gemacht haben, und neue Erkenntnisse gewonnen haben, um Therapien zu verbessern. Und das ging rein aus ... ja, klinischen Überlegungen. Dann kamen molekularbiologische Überlegungen dazu, um einen Patienten besser zu charakterisieren. Und wir bleiben trotzdem bei bestimmten Gruppen hängen, die wir nicht gesund bekommen. Das heißt, diesen steilen Anstieg in der Verbesserung von Überlebensraten geht seit ungefähr 10 Jahren in ein Plateau über. Wir können plötzlich eine bestimmte Überlebensrate nicht verbessern und wissen gar nicht, warum der eine Patient rezidiert und der andere nicht, weil unser heutiges Wissen für beide gleich ist. Das heißt, uns fehlen Informationen. Und die Informationen müssen wir aus diesen Ansätzen bekommen. Das ist die einzige Möglichkeit, indem man wirklich alle Daten von Patienten, die man hat, zusammenbringt. Indem man aus dem systembiologischer Ansatz Krankheitsmodelle entwickelt und das dann kombiniert und individuell für den einzelnen Patienten rausfindet, was ist wohl die beste Behandlung für ihn ist. Und das ist der Grund, warum das für den Kliniker eine ganz wichtige Entwicklung ist." (I18, BioMed)

These collaborations are lasting, because they are based on joint research objectives or “a common vision” (I8, IT) as one interviewee explained. Finally, the collaborating researchers are becoming more interdisciplinary-oriented and open-minded. “I think that practically all people participating in ACGT had de facto to become more multidisciplinary otherwise such a project, which is by definition a very strong multidisciplinary project, would not ever come to a successful end. I can say that we all enjoyed this opening to new areas, to new knowledge, the sharing of knowledge and interaction. It is a new window to the future somehow” (ebd.). As a result, new interdisciplinary scientific communities have emerged that are assembling around research objectives that can only be approached by interdisciplinary collaboration. As the ACGT consortium shows, because of the interdisciplinary and international approach, ICT has become an integral part of those new scientific communities such as the systems-oriented research community in oncology.

The EU commission has reacted to interdisciplinary community building by defining, for instance, the Virtual Physiological Human (VPH) as a core target of the 7th Framework Program which pursues patient-specific computer models and their applications in personalized health care (Kohl and Noble 2009). Within the frame of this program, about 30 systems-oriented research projects were funded. Their goals mainly addressed technological achievements, including data collection, management, and integration as well as processing and curation of data. Furthermore, reductionist and integrative modeling of pathophysiological processes and, finally, presentation, deployment, and end-user applications were under study.<sup>40</sup> However, references to translation were continuously included as nearly all of these projects dealt with challenges relating to patient-specific, multiscale modeling and the implementation of models and software in clinical environments. Here, simulation, data handling, scientific visualization, and community building were in the focus. Previously identified limitations in ontology annotation and inadequate tools to secure wider sharing of models and data (authentication, authorization, etc.) have also been addressed.<sup>41</sup>

Furthermore, the Virtual Physiological Human Network of Excellence (VPH NoE) was established and funded in the frame of FP 7.<sup>42</sup> The network aimed at connecting the various VPH projects and fostering the development of educational, training, and career structures for those researchers involved in VPH-related science, technology, and medicine. VPH study groups, educational meetings, and training events as well as the VPH conference series to be held every 2 years to showcase the best of VPH research were set up. In addition, the VPH NoE supported the emerging community by building up services freely available to researchers,

---

<sup>40</sup> VPH projects of FP 7 that are identified by the interviewees as follow-up projects of ACGT are, for example, p-medicine, [www.p-medicine.de](http://www.p-medicine.de), INTEGRATE, <http://www.fp7-integrate.eu/>, TUMOR, <http://tumor-project.deu/>, ContraCantrum, <http://www.contra-cancrum.eu/>. All websites accessed July 3, 2014.

<sup>41</sup> VPH Network of Excellence, Newsletter No 8, Sept 2012, VPH\_NoEnews\_N8\_34p.pdf, 10ff.

<sup>42</sup> Virtual Physiological Human Network of Excellence, <http://vph-portal.eu>. Accessed June 15, 2014.

for example, developing common standards, open source software, and freely accessible data and model repositories in the context of systems research.

To find mechanisms and strategies that enable the VPH community to continue to profit from the legacy of the VPH NoE beyond the runtime of the EU-funded network, the Virtual Physiological Human Institute for Integrative Biomedical Research (in short VPH Institute) was established and founded as an international nonprofit organization incorporated in Belgium in 2011.<sup>43</sup> Its mission is to ensure that the endeavor of the Virtual Physiological Human will be fully realized, universally adopted, and effectively used both in research and in the clinic. The VPH Institute has continued the work of the VPH NoE in many respects, including the running of the VPH conference series and the management of the VPH Portal after the VPH NoE had finished. To date, the VPH Institute represents over 67 public and private institutions active in VPH research, including many academic, clinical, and industrial key players in the area of *in silico* medicine.

To sum up, the activities of the VPH NoE show that EU science policy initiated the development of a strong interdisciplinary and Europe-wide scientific community in the field of systems biology for future research.<sup>44</sup> However, the European Union funded the network only for about five years (June 2008 to March 2013). After this funding ended, the community repositioned itself by founding the VPH Institute as an independent, nonuniversity institution. This development shows that interdisciplinary networks and research initiatives require an institutional host and permanent funding to survive. However, the universities have not yet taken up the task to fill this gap and new paths to support and foster systems research in the future are already set up that are more independent of university institutions and public funding.

### 4.3.3 *Concluding Remarks*

The broader analysis of results and outcomes of the ACGT project reveals that not all the goals of ACGT stated in the original research proposal could be realized. The reasons rest not only on a gap between too high expectations and reality of a four-year lasting research project. Rather, they refer to an underlying tension of exploring epistemic concepts and developing practice-oriented products. In particular the last aspect is closely connected to sustainability which is brought up in the interviews with regard to different objects and contexts. It thus appears important that the development of innovative products requires a research context in which concepts and practices (e.g., programming, defining parameters) are jointly developed to connect epistemic concepts with practice-oriented problems systematically.

---

<sup>43</sup>VPH Institute, [www.vph-institute.org](http://www.vph-institute.org). Accessed June 23, 2014.

<sup>44</sup>The impact of science policy on scientific developments is further discussed in Chap. 5.

The first insight of our analysis suggests that an infrastructure in a field of application needs in the first place technological sustainability. The frequently mentioned criticism regarding the Grid software shows that the technical design of the infrastructure has to provide an essential basis for sustainability. Furthermore, it is obligatory that an infrastructure in order to continue requires the conversion of a prototype into a production system that can be used by clinicians who are not familiar with high-performance computing. The ACGT platform was a prototype as the crucial steps of software engineering were still missing in terms of testing, consolidating, documenting, and certifying the software developed. The interviewees addressed the issue that university scientists are often not familiar with the systematic development of an innovative tool toward a product ready for the market. In this regard, there was uncertainty of the product maturity that could be expected at the end of the ACGT project. Other interviewees observed that academic researchers, once they have reached the proof of concept and published it, have not much interest in converting it into a production system. Here, the neglected gap between university and the market become apparent. Private companies, on the other hand, do not primarily participate in academic research to take care of marketability of research results but to understand better the trends in future research and to be involved in innovative developments. Hence, neither ACGT participants coming from academia nor industry seemed to be willing or able to be responsible for market introduction of the ICT infrastructure developed. This attitude of both academia and industry has an important impact on the translation of systems biological knowledge and tools to applied research contexts such as systems medicine.

The second insight of our analysis is that the financial sustainability of the ACGT infrastructure was not given because the technical partners were not willing to provide server capacities for an indefinite time after the ACGT project had ended. At universities, computing resources are generally limited and only used in ongoing research projects. The lack of server capacity is partially due to the Grid technology itself as the distributed servers (and technical partners) have to stay in connection to run the integrated platform. Servers where the software after the completion of a research project is hosted require continuous financial support and manpower for maintenance. Despite intense discussions on how to solve this financial problem, the ACGT consortium was not able to solve this issue.

The third insight of our analysis implies that a sustainable infrastructure requires powerful funding bodies that are able to provide a long-term perspective in terms of institutional sustainability. This is a decisive aspect for the development of systems-oriented approaches in general as new interdisciplinary scientific communities have emerged that assemble around research objectives that can only be approached by interdisciplinary collaboration. As the VPH NoE illustrates, the interdisciplinary community building is broadly funded by the European Commission. However, EU funding is time-limited and systems-oriented communities still lack institutionalization at universities. They have already reacted to this situation by founding, for example, the VPH Institute which is independent of university and public funding. Yet, it is still one of the basic challenges in systems-oriented research to find a host to institutionalize ICT infrastructures. The case of ACGT has shown that R&D projects

would not have the institutional power to build up sustainable structures. Hence, it is expected that these new forms of institutionalization may serve as a sustainable host for ICT infrastructures. Time will tell if newly founded institutions such as the VPH Institute or STaRC are able to become powerful enough to coordinate ICT in systems-oriented research, at least at the national or even European level.

The last insight is a very obvious one: an infrastructure to be consistent has to be adapted in the field of application. As already broadly discussed, the ACGT platform itself was not integrated into clinical practice until the end of the project. Hence, we show in the last section what the ACGT consortium basically expected from ICT in systems-oriented research in oncology.

#### **4.4 Impact of ICT on Systems-Oriented Research in Oncology**

Asked for the relevance of ICT, all interviewees expressed high expectations regarding the use of ICT infrastructures in systems-oriented research in oncology and in research on other diseases. In their view, ICT is indispensable because of the development towards data-intensive science and the requirement in the medical domain to translate knowledge from one research field to the other.

Infrastructures could be a breakthrough. Because something that really manages the integration of the laboratory knowledge with clinical trials and so on. The knowledge out there might have much more power than what we have now. The thing is, especially in the field of cancer, that the amount of information that we are accumulating every day is huge. But the amount of information that we can use is really ... and that we translate to the clinic safely is very little. (I9, BioMed)

Another interviewee from the biomedical domain pointed out that in many trials in the past, only clinical data were taken into consideration that were relevant in order to compare drug A versus drug B. Now, more and more data will be used for diagnosis, prognosis, and prediction of drug response and so on. By comparing them and doing experiments on them (e.g., next-generation sequencing), the amount and the complexity of data increase even more. In the laboratories, biologists are already using the results from clinical trials to set up laboratory experiments or when they investigate the functions of a gene, they routinely include experiments testing effects on drug applications as well. "These things are already happening. What it is not, a lot of times they are not happening in a structured way. Projects like ACGT help structuring the process of things that are already happening, but not in a structured way" (I9, BioMed). According to this view, ICT infrastructures provide a framework to administer large amounts of clinical and laboratory data and to create interoperability between those heterogeneous data sources.

It is one of the options to improve health care, not only cancer research, but everything. So it is one of the logical next steps that you would take if that technology that we have lying around can very much improve care and research. I think it's a good approach to test

that. We don't know it for sure. Everyone thinks that it will and everyone has a good feeling about it, but to be honest, we don't know it one hundred percent for sure. I mean there is so much data in care that you want to use in research. That is already available there. There is so much knowledge generated in research that actually has a very hard time to find its way back into care. And what all these projects try to do is to reconcile, to bring those two worlds closer together. So that researchers get more data, get more patients in their trials, and on the other hand, care providers get more direct input, assistance by mining these data to have new guidelines for treatment. They can get this immediately into the systems in form of the decision support. They also get more feedback about their personal patient if someone has done research on data and they found something weird. Then they can get feedback about it. So they have to benefit. The translational research is real. Both partners can benefit from each other and the only way in my view is to do it through ICT. (I11, IT)

The quotation stresses the impact of ICT on translational research by aligning clinical and laboratory research and designing “the structured way” (previous quotation) how clinical practice and laboratory research relate to each other. However, the decisive point is that some procedures and tasks would not be performed without ICT support. For example, the semantic and syntactic integration of different data types based on corresponding standards is regarded as necessary (see Sect. 4.1.3). ICT infrastructures can therefore be described as catalysts of doing translational research. Quite similar, the term “breakthrough” was used in the interviews to describe the impact of ICT on translational processes from the laboratory to the clinic and vice versa (see previous quotation).

However, the description of ICT as a catalyst was only one picture in the interviews. An opposing picture is outlined by the following citation.

To translate and do it safely, you do need a huge process that can be accelerated if you have good integration of data and you are using the standards and so on. [...] It could be a huge facilitator if you really have a good platform. The discovery might come anyway, but it takes ten years instead of one year if you don't have a good infrastructure. (I9, BioMed)

Here, the term used, “facilitator” (previous citation), indicates that ICT should ease and accelerate research activities. However, according to this interpretation, ICT is not indispensable for translational research as “the discovery might come anyway” (previous citation). Hence, the assigned tasks of ICT infrastructures (e.g., providing access and making data shareable) are not regarded as being part of the original research process. ICT infrastructures rather appear as a data management system and in this sense as a service facility.

As a result, there are two opposing concepts of how ICT functions in systems medical research outlined in the interviews. The first picture (ICT as a catalyst) reflects our second hypothesis outlined in the beginning of this chapter that the application of ICT enables doing systems-oriented research as some research activities would not be possible without ICT support. The second picture (ICT as a facilitator) mirrors the more popular understanding, namely that the ICT infrastructure as a service facility are not taking part in research processes. We discuss in the last section, what kind of function and role ICT infrastructures may in fact play in systems-oriented research in the future.

### 4.4.1 Conclusion

The ACGT project had the ambitious goal of designing an ICT infrastructure in support of systems-oriented approaches in oncology and implementing it into the emerging scientific community. The integrated platform aimed at offering tools and techniques for the distributed mining of autonomous data repositories and extraction of knowledge by knowledge discovery services.

However, as discussed in Sect. 4.1, the development of the ACGT infrastructure was accompanied by considerable challenges coming from different angles. (1) The overarching task of data integration had to be tackled by considering syntax, semantics, and data acquisition contexts; (2) many technological problems occurred and had to be solved on an ad hoc basis in the course of its development as basic standards for integration processes did not yet prevail; and (3) the necessity to work together within different disciplines required not only individual skills but also elaborated strategies of project management to arrange interdisciplinary collaboration.

As shown in Sect. 4.3, the ACGT project was a pioneering project and the interviewed consortium members often referred to the status of ACGT as an R&D project. It was stressed that it represented one of the first approaches as to how Grid technology could be applied to support and facilitate medical research. Therefore, the consortium had to start conceptually from scratch. The initial goal was to explore whether the Grid technology would be adaptive to the needs and demands of the systems-oriented research community in oncology. At the same time, the interviewees agreed that an infrastructure that can be used in clinical practice requires much more effort in terms of financial support and time to be invested into software engineering. However, the implementation to the clinic was far from realization as the ACGT participants—neither from academia nor from industry—were not able to take over the responsibility to steer this process of marketability in the lifespan of the project. Hence, many of them in the beginning did not believe that it was possible to develop an ICT infrastructure that would be approved by the oncological research community after four years of research.

What was finally developed was a core set of technological components. They were assembled to develop an architectural prototype that was presented as an end-to-end demonstration at the final review meeting. However, the Grid technology was criticized as too complex and too immature by some interviewees and it was consequently replaced in follow-up projects by Cloud technology. In addition to the technological tools and services developed in the course of the project, many interviewees referred to more indirect outcomes. They had a lot to say about gaining experience in the emerging field of ICT in the medical domain. In this context, the ACGT project was often evaluated as the beginning of making a career in interdisciplinary research merging ICT, medicine, and systems biology and as the beginning of lasting interdisciplinary networking. These indirect benefits were even more positively assessed compared to technological achievements as impact on research developments was rather connected to experience and networking than to technological innovation. In particular, gaining experience provides the basis for long-acting achievements in a research field which was often connected to reach impact on

research developments. In this context, the work on semantics and in particular on metadata definitions for describing data and the capabilities of tools was highlighted. This work of the ACGT consortium was valued as very important because it influences standardization processes in the research field. They were acknowledged as indispensable prerequisites for the adaptation of ICT infrastructures into clinical practice. By referring to the submission of the ACGT Master Ontology to the Open Biomedical Ontology Foundry, it was claimed in the interviews that the initial foundational work was already done in the domain of metadata definitions that has been capitalized in follow-up projects. However, the ACGT MO was not approved by the quality assurance of the OBO Foundry and, as already broadly discussed, the ACGT platform itself was not integrated into clinical practice.

The OBO Foundry is a good example of standardization efforts in systems medicine. These processes have a crucial impact on the coordination of systems medicine in an ICT environment. As described in Sect. 4.1.3, technological tools and services for the overarching task of data integration are consistently developed on the basis of ICT standards: the ICT formats of existing tools work as standards for the new ones. Because the tools based on ICT standards address syntactic as well as semantic integration, ICT deploys the standards for data storing and processing as well as the standards for data quality, annotation, and exchange. ICT-based standards therefore define what counts as reliable and valid data in the research process. Hence, ICT environments collect and integrate not only data on a technological level, they construct at the same time the data used in system-oriented research by assigning significance and meaning to them.

However, it seems as if the interviewees were not conscious of the ICT's profound impact on systems-oriented research. Unambiguously, they appreciated ICT infrastructures as data management systems providing access to and integration of large heterogeneous data stocks. Responsibility for those activities associated with standardization, integration, and management of data have therefore been ascribed to the scope of the ICT infrastructure and not to research institutions or individual scientists any more. In addition, the advantages of using ICT were regarded as being easy to connect and collaborate within the emerging scientific community of international and interdisciplinary range. In line with this perception is the popular picture of the ICT infrastructure as a facilitator or, in other words, as a service facility. Integration, access, and sharing of data are assigned to ICT infrastructures. These tasks are defined in technological terms assessed as general functions of a management system. The characterization of ICT as a facilitator corresponds to the emphasis on standardization and the categorical division of data into structure and content (see Sect. 4.1.3). Although data are split up into one part that is designated for scientific investigation and another part that contains purely technical information, the corresponding responsibilities of collecting and managing data are now separated from researching data. Accordingly, the ICT infrastructure needs to be stable, static, and enduring, whereas the actual use of stored data by scientists is claimed to be dynamic, creative, and proceeding. Not surprisingly, this picture of ICT infrastructures as service facilities was commonly used in the interviews as it is the most popular picture of ICT in science (e.g., Nyrönen et al. 2012).

However, we assumed that, looking from an in-depth perspective of a case study, understanding and modeling of biomedical systems are deeply shaped by ICT and their underlying design and conceptualization. Therefore, we expected to find evidence for this hypothesis as well. In fact, our hypothesis was corroborated by a second picture found in the interviews. Some of the interviewees characterized ICT infrastructures as catalysts for shaping and transforming systems-oriented research. The decisive point of this argument is that some research activities would not be performed without ICT support. Examples given in the interviews were tasks and processes regarding interoperability between heterogeneous data sources or knowledge discovery workflows (e.g., data mining services, the workflow editor, or the oncosimulator). The detailed analysis of the oncosimulator (see Sect. 4.2) has shown that the development of such a systems-oriented research tool in an ICT environment is based on many indicative decisions (e.g., using a top-down approach, prioritizing particular mathematics) paving the way intrinsically to integrate ICT in systems-oriented research. However, insights into the oncosimulator's conceptual grounding have revealed that ICT not only enables but also restricts doing research at the same time. The exclusive use of mathematically compatible parameters indicated, for example, those kinds of restrictions.

To explore this transition of scientific and technological processes and mechanisms, it may be useful to look at it from a sociological perspective. The Actor-Network-Theory (ANT)<sup>45</sup> conceptualizes society as a completely interwoven sociotechnical web in which both parts, the social and the technical, influence each other mutually. This principle of symmetry between technology and humans rejects both technological determinism and social determinism and analyzes the mechanism of interactions in human–technological networks. Such sociotechnical networks can only exist when human and nonhuman actors (actants)<sup>46</sup> are permanently connected. They are therefore semiotically defined by how they act and are acted on in the networks of practices. The humans and nonhumans have a certain role and perform certain tasks within the network while delegating other roles and tasks to other actants (Latour 1992). In other words, the actants relate their roles and agency to each other. After the network is finally coordinated, it exists as an independent functional unit having agency on its own.

According to the ANT's perspective, ICT infrastructures in systems medicine operate as a nonhuman actant integrated together *inter alia* with scientists, scientific organizations, and funding organizations into a sociotechnical network. Technical objects such as an ICT infrastructure have a mediating role in the development of a network as they build, maintain, and stabilize the relations between different actants of all types and sizes, whether human or nonhuman. This means that they embody

---

<sup>45</sup> ANT was mainly developed by Bruno Latour (1987, 2005), Michel Callon (1986) and John Law (1987; Law and Hassard 1999). For a collection of original ANT papers in German see Belliger and Krieger (2006).

<sup>46</sup> To stress the interaction of nonhumans and humans, one term for both is preferred. According to ANT-terminology, they are defined as actants (e.g., Latour 1996). Nonhuman actors may be, for instance, technical artifacts, laboratories, or companies (Callon 1992, 73).

and measure relations between different actants at the same time. However, this does not mean that they are not actants of the sociotechnical network themselves (Akrich 1992, 205f.).

We argue that the ICT's role in the network is systematically to align the different data acquisition contexts with research activities (see Sect. 4.1.3). Therefore, the ICT infrastructure acts as a bridge between laboratory and clinic, molecular data and clinical data, as well as data acquisition and data interpretation. In the end, the interviewees expected that the ICT infrastructures have the potential to pave the way towards systems and personalized medicine; ICT infrastructures are, as one interviewee put it, "a new window to the future." From this it follows that the departure into the new era of systems medicine basically relies on the use of ICT technology: ICT infrastructures are promised to become new powerful actors (or actants, respectively) in upcoming networks in systems medicine.

However, the analysis in Sect. 4.3.2 has shown that ICT infrastructures require a suitable frame to be successfully integrated into systems-oriented research. In the interviews, different aspects of sustainability were outlined as being necessary to maintain an infrastructure over time. With the termination of the ACGT platform in mind, the interviewees drew on new approaches and concepts of how to tackle technological, financial, and social sustainability of ICT infrastructures for systems-oriented research. In addition to an appropriate technological design, it is obligatory that for an ICT platform to continue it has continuous financial support and manpower for maintenance. Providing financing on an ongoing basis is a sensitive issue not yet solved by the European Commission. Even if the use and development of research infrastructures is an overall objective in the 7th EU Framework Program and in its follow-up program Horizon 2020 (see Sect. 5.1), the EU commission usually funds research projects in the start-up phase of ICT infrastructures only. Hence, the original goal of the ACGT consortium to implement the ACGT platform into clinical practice was condemned to failure right from the beginning because of the lack of sustainability afterwards.

However, participating in R&D projects was highly attractive for the interviewees. They valued R&D projects as gateways for setting trends in research and for inter- and transdisciplinary networking. These highly appreciated advantages of participating in EU-funded projects are used to strike a new path of independence from EU money: to secure long-term funding of ICT infrastructures, new institutionalized frames such as a business plan for the Study, Trial and Research Center, STaRC, or the new foundation of the VPH Institute as a nonprofit organization are about to be realized. As these spin-offs are still very young, the connection to EU-funded research is very close. STaRC, for example, is intertwined with the EU project p-medicine and the VPH Institute can be regarded as follow-up of the VPH Network of Excellence. Within these emerging institutions, the key actors are scientific managers often trained and networked in EU-funded research projects. They need to be inter- and transdisciplinary oriented the more that science, medicine, funding, industry, and politics merge. Of course, these scientific managers are still interested in doing cutting-edge research and realizing systems or personalized medicine. However, by stepping outside the academic world, they have to consider more actors and interests coming from different grounds.

## References

- Abu-Asab MS, Abu-Asab N, Loffredo CA, Clarke R, Amri H (2013) Identifying early events of gene expression in breast cancer with systems biology phylogenetics. *Cytogenet Genome Res* 139(3):206–214
- ACGT (2005) Annex 1—Description of Work, Proposal, unpublished work
- ACGT (2009) Demonstrator specifications, deliverable 13.4, unpublished work
- Akrich M (1992) The description of technical objects. In: Bijker W, Law J (eds) *Shaping technology/building society: studies in sociotechnical change*. MIT Press, Cambridge, pp 205–224
- Ankeny RA, Leonelli S (2011) What is so special about model organisms? *Stud Hist Philos Sci A* 42(2):313–323
- Auffray C, Chen Z, Hood L (2009) Systems medicine: the future of medical genomics and healthcare. *Genome Med* 1(1):2
- Belliger A, Krieger D (2006) ANThology. Ein einführendes Handbuch zur Akteur–Netzwerk–Theorie. transcript, Bielefeld
- Bernsen HJ, van der Kogel AJ (1999) Antiangiogenic therapy in brain tumor models. *J Neurooncol* 45(3):247–255
- Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C et al (2001) Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nat Genet* 29(4):365–371
- Brochhausen M, Blobel B (2011) Architectural approach for providing relations in biomedical terminologies and ontologies. *Stud Health Technol Inform* 169:739–743
- Brochhausen M, Spear AD, Cocos C, Weiler G, Martín L, Anguita A et al (2011) The ACGT Master Ontology and its applications—towards an ontology-driven cancer research and management system. *J Biomed Inform* 44(1):8–25
- Bucur A, Rüping S, Sengstag T, Sfakianakis S, Tsiknakis M, Wegener D (2011) The ACGT project in retrospect: lessons learned and future outlook. *Proc Comput Sc* 4:1119–1128
- Burgoon LD (2007) Clearing the standards landscape: the semantics of terminology and their impact on toxicogenomics. *Toxicol Sci* 99(2):403–412
- Burkhardt H, Smith B (1991) *Handbook of metaphysics and ontology*. Philosophia, Munich
- Callon M (1986) Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay. In: Law J (ed) *Power, action and belief: a new sociology of knowledge*. Routledge & Kegan Paul, London, pp 196–233
- Callon M (1992) The dynamics of techno-economic networks. In: Coombs R, Saviotti P, Walsh V (eds) *Technological change and company strategies. Economic and sociological perspectives*. Academic, London, pp 72–102
- Coebergh JW, van Veen EB, Vandenbroucke JP, van Diest P, Oosterhuis W (2006) One-time general consent for research on biological samples: opt out system for patients is optimal and endorsed in many countries. *BMJ* 332(7542):665
- Deisboeck TS, Zhang L, Yoon J, Costa J (2009) In silico cancer modeling: is it ready for prime time? *Nat Clin Pract Oncol* 6(1):34–42
- Desmedt C, Di Leo A, de Azambuja E, Larsimont D, Haibe-Kains B, Selleslags J et al (2011) Multifactorial approach to predicting resistance to anthracyclines. *J Clin Oncol* 29(12):1578–1586
- Dionysiou DD, Stamatakos GS, Uzunoglu NK, Nikita KS, Marioli A (2004) A four-dimensional simulation model of tumour response to radiotherapy in vivo: parametric validation considering radiosensitivity, genetic profile and fractionation. *J Theor Biol* 230(1):1–20
- Edwards PN, Mayernik MS, Batcheller AL, Bowker GC, Borgman CL (2011) Science friction: data, metadata, and collaboration. *Soc Stud Sci* 41(5):667–690
- Field D, Sansone SA, Collis A, Booth T, Dukes P, Gregurick SK et al (2009) Megascience. ‘Omics’ data sharing. *Science* 326(5950):234–236
- Forgó N, Kollek R, Arning M, Kruegel T, Petersen I (2010) Ethical and legal requirements for transnational genetic research. C.H. Beck, Munich

- García-Sancho M (2012) From the genetic to the computer program the historicity of 'data' and 'computation' in the investigations on the nematode worm *C. elegans* (1963–1998). *Stud Hist Philos Biol Biomed Sci* 43(1):16–28
- Graf N, Hoppe A (2006) What are the expectations of a clinician from in silico oncology? 2nd international advanced research workshop on in silico oncology (IARWISO). Kolympari, Chania, Greece
- Graf N, Hoppe A, Georgiadi E, Belleman R, Desmedt C, Dionysiou D et al (2009) 'In silico' oncology for clinical decision making in the context of nephroblastoma. *Klin Padiatr* 221(3):141–149
- Gramelsberger G (2013) Simulation and systems understanding. In: Andersen H, Dieks D, Gonzalez WJ, Uebel T, Wheeler G (eds) *New challenges to philosophy of science*. Springer, Dordrecht, pp 151–161
- Green S, Wolkenhauer O (2012) Integration in action. *EMBO Rep* 13(9):769–771
- Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144(5):646–674
- Hanseth O, Monteiro E, Hatling M (1996) Developing information infrastructure: the tension between standardization and flexibility. *Sci Technol Hum Values* 21(4):407–426
- Kitchin R (2014) Big Data, new epistemologies and paradigm shifts. *Big Data & Society* 1. doi:[10.1177/2053951714528481](https://doi.org/10.1177/2053951714528481)
- Kohl P, Noble D (2009) Systems biology and the virtual physiological human. *Mol Syst Biol* 5:292
- Kollek R (2009) Informed consent. Comment on article 6. In: Ten Have HAMJ, Jean MS (eds) *Universal declaration on bioethics and human rights. Background, principles and application*. UNESCO Publishing, Paris, pp 123–138
- Kuiper RA, Schellens JH, Blijham GH, Beijnen JH, Voest EE (1998) Clinical research on antiangiogenic therapy. *Pharmacol Res* 37(1):1–16
- Kyriazis D, Menychtas A, Dionysiou D, Stamatakos G, Varvarigou T (eds) (2008) *Clinical trial simulation in grid environments*. Conf. Proc. 8th IEEE International Conference on BioInformatics and BioEngineering (BIBE). Athens, 8–10 October 2008. 6 p
- Latour B (1987) *Science in action: how to follow scientists and engineers through society*. Open University Press, Milton Keynes
- Latour B (1992) Where are the missing masses? The sociology of a few mundane artifacts. In: Bijker W, Law J (eds) *Shaping technology/building society: studies in sociotechnical change*. MIT Press, Cambridge, pp 225–259
- Latour B (1996) On actor-network theory. A few clarifications. *Soziale Welt* 47(4):369–382
- Latour B (2005) *Reassembling the social: an introduction to actor-network-theory*. Oxford University Press, Oxford
- Law J (1987) Technology and heterogeneous engineering: the case of portuguese expansion. In: Bijker W, Law J (eds) *The social construction of technological systems: new directions in the sociology and history of technology*. MIT Press, Cambridge, pp 111–134
- Law J, Hassard J (eds) (1999) *Actor network theory and after*. Blackwell and the Sociological Review, Oxford
- Lee JK, Coutant C, Kim YC, Qi Y, Theodorescu D, Symmans WF et al (2010) Prospective comparison of clinical and genomic multivariate predictors of response to neoadjuvant chemotherapy in breast cancer. *Clin Cancer Res* 16(2):711–718
- Leonelli S (2010) Packaging data for re-use: databases in model organism biology. In: Howlett P, Morgan MS (eds) *How well do facts travel? The dissemination of reliable knowledge*. Cambridge University Press, Cambridge, pp 325–348
- Leonelli S (2012) Introduction: making sense of data-driven research in the biological and biomedical sciences. *Stud Hist Philos Biol Biomed Sci* 43(1):1–3
- Leonelli S (2014) What difference does quantity make? On the epistemology of big data in biology. *Big Data & Society* 1. doi:[10.1177/2053951714534395](https://doi.org/10.1177/2053951714534395)
- Leonelli S, Ankeny RA (2012) Re-thinking organisms: the impact of databases on model organism biology. *Stud Hist Philos Biol Biomed Sci* 43(1):29–36
- Leonelli S, Diehl AD, Christie KR, Harris MA, Lomax J (2011) How the gene ontology evolves. *BMC Bioinformatics* 12:325

- Lin B, White JT, Lu W, Xie T, Utleg AG, Yan X et al (2005) Evidence for the presence of disease-perturbed networks in prostate cancer cells by genomic and proteomic analyses: a systems approach to disease. *Cancer Res* 65(8):3081–3091
- Mayer-Schonberger V, Cukier K (2013) *Big data: a revolution that will transform how we live, work and think*. John Murray, London
- Meier S, Gehring C (2008) A guide to the integrated application of online data mining tools for the inference of gene functions at the systems level. *Biotechnol J* 3(11):1375–1387
- Meuser M, Nagel U (1991) ExpertInneninterviews—vielfach erprobt, wenig bedacht. Ein Beitrag zur qualitativen Methodendiskussion. In: Garz D, Kraimer K (eds) *Qualitativ-empirische Sozialforschung. Konzepte, Methoden, Analysen*. Westdeutscher Verlag, Opladen, pp 441–471
- Michelson S, Sehgal A, Friedrich C (2006) In silico prediction of clinical efficacy. *Curr Opin Biotechnol* 17(6):666–670
- Nyrönen TH, Laitinen J, Tourunen O, Sternkopf D, Laurikainen R, Öster P et al (2012) Delivering ICT infrastructure for biomedical research. *Proc. WICSA/ECSA Companion Volume*. pp 37–44
- O'Malley MA, Soyer OS (2012) The roles of integration in molecular systems biology. *Stud Hist Philos Biol Biomed Sci* 43(1):58–68
- Rakyan VK, Down TA, Balding DJ, Beck S (2011) Epigenome-wide association studies for common human diseases. *Nat Rev Genet* 12(8):529–541
- Rubin DL, Shah NH, Noy NF (2008) Biomedical ontologies: a functional perspective. *Brief Bioinform* 9(1):75–90
- Sanga S, Frieboes HB, Zheng X, Gatenby R, Bearer EL, Cristini V (2007) Predictive oncology: a review of multidisciplinary, multiscale in silico modeling linking phenotype, morphology and growth. *Neuroimage* 37(Suppl 1):S120–S134
- Stamatakos G (2011) In silico oncology part I: clinically oriented cancer multilevel modeling based on discrete event simulation. In: Deisboeck T, Stamatakos G (eds) *Multiscale cancer modelling*. CRC, Boca Raton, pp 407–436
- Stamatakos GS, Antipas VP, Uzunoglu NK (2006a) A spatiotemporal, patient individualized simulation model of solid tumor response to chemotherapy in vivo. The paradigm of glioblastoma multiforme treated by temozolomide. *IEEE Trans Biomed Eng* 53(8):1467–1477
- Stamatakos GS, Antipas VP, Uzunoglu NK, Dale RG (2006b) A four dimensional computer simulation model of the in vivo response to radiotherapy of glioblastoma multiforme. Studies on the effect of clonogenic cell density. *Br J Radiol* 79:389–400
- Stamatakos GS, Dionysiou DD, Graf NM, Sofra NA, Desmedt C, Hoppe A et al (2007) The “Oncosimulator”. A multilevel, clinically oriented simulation system of tumor growth and organism response to therapeutic schemes. Towards the clinical evaluation of in silico oncology. *Conf. Proc. IEEE Eng Med Biol Soc*: pp 6629–6632
- Stamatakos GS, Dionysiou DD, Zacharaki EI, Mouravliansky NA, Nikita KS, Uzunoglu NK (2002) In silico radiation oncology: combining novel simulation algorithms with current visualization techniques. *Conf. Proc. IEEE Special Issue Bioinformatics Adv Challenges* 90(11): 1764–1777
- Stamatakos GS, Kolokotroni E, Dionysiou D, Veith C, Kim YJ, Franz A et al (2013) In silico oncology: Exploiting clinical studies to clinically adapt and validate multiscale oncosimulators. *Conf Proc IEEE Eng Med Biol Soc*: 5545–5549
- Sujanski W (2001) Heterogeneous database integration in biomedicine. *J Biomed Inform* 34(4):285–298
- Swertz MA, Jansen RC (2007) Beyond standardization: dynamic software infrastructures for systems biology. *Nat Rev Genet* 8(3):235–243
- Symmans WF, Hatzis C, Sotiriou C, Andre F, Peintinger F, Regitnig P et al (2010) Genomic index of sensitivity to endocrine therapy for breast cancer. *J Clin Oncol* 28(27):4111–4119
- Thomas G (2011) A typology for the case study in social science following a review of definition, discourse and structure. *Qual Inq* 17(6):511–521
- Tomita M (2001) Whole-cell simulation. A grand challenge of the 21st century. *Trends Biotechnol* 19(6):205–210

- Tsiknakis M, Kafetzopoulos D, Potamias G, Analyti A, Marias K, Manganas A (2006) Building a European biomedical grid on cancer: the ACGT integrated project. *Stud Health Technol Inform* 120:247–258
- Van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M et al (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415(6871):530–536
- Werner-Wasik M, Scott CB, Nelson DF, Gaspar LE, Murray KJ, Fischbach JA et al (1996) Final report of a phase I/II trial of hyperfractionated and accelerated hyperfractionated radiation therapy with carmustine for adults with supratentorial malignant gliomas. Radiation Therapy Oncology Group Study 83-02. *Cancer* 77(8):1535–1543
- Wierling C, Herwig R, Lehrach H (2007) Resources, standards and tools for systems biology. *Brief Funct Genomic Proteomic* 6(3):240–251
- Wolkenhauer O, Green S (2013) The search for organizing principles as a cure against reductionism in systems medicine. *FEBS J*. doi:[10.1111/febs.12311](https://doi.org/10.1111/febs.12311)
- Wolkenhauer O, Auffray C, Jaster R, Steinhoff G, Dammann O (2013) The road from systems biology to systems medicine. *Pediatr Res* 73(4 Pt 2):502–507
- Yin RK (2009) *Case study research: design and methods*, 4th edn. SAGE, Thousand Oaks