Digital Scientific Practice in Systems Medicine

Europäische Vergesellschaftungsprozesse im Spannungsfeld von System- und Sozialintegration

Imme Petersen[1], Regine Kollek[2]

Beitrag zur Ad-hoc-Gruppe »Soziologische Perspektiven auf die Cloud« – organisiert von Michael Eggert und Daniel Kerpen

Introduction

Evolving from systems biology, systems medicine tries to grasp the interactions between all components of health and disease to create a basis for innovative therapies and preventive measures (e.g. Fedoroff, Gostin 2009). To illuminate complex physiological and pathological processes, high-throughput technologies for molecular profiling are employed to produce large stocks of data on genomes, transcriptomes and proteomes. The variety of such data and their large volume require new approaches in data handling and processing. Compared to less data-intensive scientific practices in the life sciences, high capacity databases and infrastructures had to be developed to digitally store the different data types and to link separate databases at different geographical sites. High throughput production and digitalisation of scientific data have an enormous impact on how systems medical research is done today and what it may be able to achieve. Therefore it is critical to thoroughly understand the impact of the information and communication technology (ICT) employed and to assess the consequences it may have for the process of knowledge production. In other words: Is the impact of ICT simply a matter of speed and scale, as they store, process and analyse large amounts of data in very short time? Or does ICT make more fundamental changes with regard to the scientific practices of how knowledge is produced, organised and utilised?

Based on the concept of scientific practice introduced by Andrew Pickering (1992), Michael Lynch (1993), Joseph Rouse (2002), Karin Knorr-Cetina (1999) and others, this paper focuses on different factors and mechanisms that play a role in initiating and supporting the use of ICT in

[1] Research Centre for Biotechnology, Society and the Environment (FSP BIOGUM), Universität Hamburg, Lottestrasse 55, 22529 Hamburg, email: imme.petersen@uni-hamburg.de

[2] Research Centre for Biotechnology, Society and the Environment (FSP BIOGUM), Universität Hamburg, Lottestrasse 55, 22529 Hamburg, email: kollek@uni-hamburg.de

daily scientific practice. By exploring the challenges generated by the concepts and methods of ICT-based research, we will argue that ICT is not only a science-supporting technology, but is deeply engraved in its practices of knowledge generation. Findings supporting our argument are derived from an empirical case study in which we analysed the challenges of organising and doing systems medical research in an ICT environment as well as the strategies of interlinking conceptual and material resources coming from both domains. The case under study is a research project in which an integrated European ICT infrastructure was designed and developed in order to support the systems medical research community in oncology.

In the following, the specific ways that systems medicine research produces, stores, and manages data in an ICT environment will be traced. In particular, how ICT modify the way data are perceived epistemically and what consequences arise from such ICT-induced epistemic shifts will be outlined. As the primary objective of ICT is to integrate and harmonise data, the standardisation processes will in particular be relevant for our reflexion regarding the technology's role in scientific practice.

Information and Communication Technology in Systems Medicine

High-throughput technologies for molecular profiling - first used on a broader scale in the Human Genome Project - have produced large stocks of data of different types, often referred to as omics data. Omics data are generated from genomes, expressed DNA-sequences (transcriptomes), cellular proteins (proteomes) or metabolic pathways (metabolomes) in order to understand the mechanisms underlying health and disease and to identify and or even prevent potential pathologies as well as disease development and its progression. How components of cells and organisms in pathological processes relate to and interact with each other has been increasingly formalised in mathematical models. Their goal is to systematically explore such interactions by simulating metabolic pathways, virtual cells or organisms in the computer (Janes and Yaffe 2008). These activities were stimulated by the vision of finally being able to make individualised diagnoses, provide effective treatments for all kinds of diseases available and inform the public about how best to preserve their health and well-being (Weston, Hood 2004).

However, the plethora of data on (individual) genomes, transcriptomes and proteomes brought new urgency to the problem of how to deal with the large amount of heterogeneous data usually stored in separate databases at different geographical sites. Indeed, integration of patient-related data coming from such autonomous, distributed databases raises questions concerning data protection and audited data access (e.g. Forgó et al. 2010). Further problems have been identified with regard to data integration. For one, there are problems related to the fact that the data, coming from different data sources, significantly vary with regard to the circumstances they were collected (e.g. clinical or research context, history of data collection, local institution, national law). As well, the terminology and classification of data often differ between different databases (e.g. Meier, Gehring 2008). This situation, which creates numerous hurdles or even an inability to share data, is considered to be a major obstacle to the research process in systems medicine (e.g. Tsiknakis et al. 2006: 248). Even integrating data created by different technologies within a single

laboratory seems to be difficult; thus efficient research collaborations within the wider scientific community are demanding or problematic (Sweertz, Jansen 2007).

The increasing challenges of storing and sharing omics data gave rise to the development of new methodological and conceptual approaches. To systematise, integrate and share data from genomic research of a given domain (e.g. disease), computational-based databases and infrastructures were developed in the late 1990s and early 2000s. The use of digital technology has clear pragmatic advantages for scientific practice: It provides a comprehensive data management system, it enables secure access to databases on a global basis, it facilitates simultaneous access to various types of information for comparison, and it usually provides open access to all interested researchers (Leonelli, Ankeni 2012: 31). As a result, researchers who use ICT infrastructures in their scientific practice have the ability to access more and more data from available sources and to easily connect and collaborate within the emerging interdisciplinary and internationally growing communities of systems medicine (Leonelli 2014).

Empirical Approach

The work presented here was conducted in the context of the research project "Towards a holistic conception of life? Epistemic presumptions and socio-cultural implications of systems biology" (THCL), which was carried-out by the research group composed of Martin Döring, Anne Brüninghaus, Imme Petersen and Regine Kollek (Prinicipal Investigator) at the Research Centre for Biotechnology, Society and the Environment (FSP BIOGUM) at Universität Hamburg. The research was funded by the German Federal Ministry of Education and Research (BMBF) (project no.: 01GP0904) from May 2010 to July 2014. Based on an empirical social study of science, the project investigated the current systems approach in the life sciences. In more detail, the empirical analysis explored the epistemic preconditions, scientific practices, innovative potentials and policy implications of emerging and expanding concepts and practices of systems biology (Döring et al. forthcoming).

To explore the scientific practices in an ICT environment, one of the empirical studies of the project focused on the respective infrastructural requirements of systems biological and medical research. Imme Petersen and Regine Kollek analysed the conception and realization of an ICT infrastructure in the domain of cancer research. The case under study was the research project "Advancing clinico-genomic trials on cancer: open Grid services for improving medical knowledge discovery (ACGT)" funded by the 6th Framework Program of the European Commission (FP6/2004/IST-026996).[3] From February 2006 until July 2010, 26 research groups from twelve Eu-

─────────

[3] By intensely looking at an individual case, conclusions can only be drawn about that case in its specific context. From this 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).

ropean countries and Japan designed and developed an integrated technological platform in support of postgenomic, multi-centric clinical trials targeting two major cancer diseases, namely breast cancer and paediatric nephroblastoma.[4]

The ultimate initial aims of the ACGT consortium were to design experiments for obtaining coherent and consistent medical and biological data; to develop methods for integrating heterogeneous data (e.g. genomic, clinical); to develop methods for selection, checking, cleaning, and pre-processing of combined genomic-medical data; and to incorporate collaborative approaches to data analysis (ACGT 2005: 9).[5] To address these goals, different ICT tools and services were developed and implemented into the ACGT infrastructure.

The main technical components are:

(1) biomedical technology Grid layer: The Grid technology comprises the basic technology for the scheduling and brokering of resources;

(2) distributed data access and applications. A set of software services based on web services provide uniform data access to distributed and heterogeneous data sources, i.e. clinical data, eHealth records, microarrays, SNP data, etc;

(3) ontologies and semantic mediation tools. Formalised knowledge representations facilitate semantic data integration as well as annotation and data analysis of large-scale biomedical data;

(4) clinical trial management system. The clinical trial builder based on an ontology-driven software aims to help easily set up new clinico-genomic trials;

(5) technologies and tools for in-silico-oncology. The oncosimulator models tumour growth and therapy response in silico;

(6) Grid-enabled application layer. The data-mining Grid services support and improve complex knowledge discovery processes and knowledge extraction operations;

(7) the integrated ACGT architecture. Integration of applications requires a composite service that orchestrates other services in order to interoperate in a workflow (ACGT 2005, 11f.).

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. Furthermore, tools for the analysis of such integrated data were provided and supported by discovery-driven analytical workflows. As well, these research activities were in compliance with existing ethical and legal regulations (ACGT 2005: 10; Bucur et al. 2011: 1120). However, several months after the ACGT project was completed, the integral infrastructure was shut down due to missing server capacities and an overall lack of financial sustainability (see Chapter 4.3 in Döring et al. forthcoming).

——————

[4] Postgenomic clinical research investigates the role of the genome in the development and progression of disease, for example, which genes are active at particular times and under different environment conditions (gene expression).

[5] As the authors were part of the ACGT consortium, we had access to unpublished working papers, e.g. ACGT. 2005: Annex 1 – Description of Work, Proposal.

After the project was finished, we conducted guided interviews with participants of the consortium. To select interview partners, we identified the most relevant actors within the ACGT consortium keeping the consortium and the project running. In order to choose the relevant scientists, we assumed that the ACGT consortium was a network of actors working together on the joint task to develop an integrative ICT infrastructure. Accordingly, the most relevant actors were the ones working most intensely in cooperation with other ACGT members. As publications resulting from such large research projects are usually based on joint work, we conducted a bibliometric analysis of internal (deliverables) and external publications (peer-reviewed articles, books, conference proceedings). Our data cluster conjointly comprised co-authorship and cooperation within the ACGT consortium. From this we contacted the 20 first most active project participants. 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. It 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 in building the ICT infrastructure (in particular tasks and challenges); (3) judgments regarding the project outcome and science policy on supporting ICT in science; and (4) judgments regarding the anticipated profit of ICT for systems medical research in oncology and systems biology. The interviews were digitally recorded, anonymised and transcribed literally. 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 characterising the topics that were jointly discussed in the interviews (Meuser, Nagel 1991). Below, the interview citations are characterised by the professional background of the interviewee.

In addition to the interview material, we analysed 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.

Functions and Meanings of ICT in Systems Medicine

The primary goal of the ACGT consortium was to build an ICT infrastructure that provides data management for heterogeneous data stemming from different sources ranging from structural analysis of molecules (DNA, protein), pathology and other laboratory examinations, physiological and clinical data findings on case report forms (e.g. symptoms, histol ogy, administered treatment, treatment response) to imaging data (MRI, CT, ultrasound). In addition to this, many external sources of data and knowledge about gene and protein sequences, pathways, genomic variations, microarrays experiments, etc. are systematically collected and used. Generally, all the collected data are immediately stored in digital databases (Bucur et al. 2011: 1120) because digital

systems are best suited to store large amounts of data and information (Ankeny, Leonelli 2011). However, many obstacles still occur, as one of the interviewees of our case study pointed out:

> *"Due to the developments of the times, 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 realised 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 technological tools that need to exist to analyse these data. (Support for) more efficient ways of distributing teams that by nature are involved in such interdisciplinary types of research, clinicians like molecular biology, computer scientists, etc. is needed. So there are a lot of inefficiencies in the processes of semantics, harmonising the data and the representation of the data; developing shared tools therefore supports this concept of an open-source sharing of tools, avoiding reinventing the wheel, etc. Every specific lab investing in developing their own computational solutions and platforms. And therefore, the vision and the ultimate objective were 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 organisation supported by an information technology solution"* (Interview 7, IT)

By emphasising the inefficiencies and obstacles to bring large stocks of heterogeneous data together and to develop shared tools to analyse such data, the interviewee above addresses insufficient integration of data and tools as a key problem in systems medicine today. In the literature on living systems, data integration widely refers to the process of making different data types comparable from a variety of different data sources (e.g. Philippi, Köhler 2006; Schmid, Blank 2010). In this process of theorising and modelling data sets, the data are quantified and accurately cleaned as well as edited in a way so that it can be re-used and re-analysed in novel ways (Ge et al. 2003; Lenzerini 2002). Here, data standardisation comes to the fore in order to finally integrate data:

> *"Therefore, data have to be standardised or homogenised in (order to move data into) large data sets for data processing. Because studies or domains speak their own language, data can look quite different. Standardising all data was the big challenge."* (Interview 1, IT)

Classical approaches to solve this challenge focus on syntactic interoperability, which basically means that two or more databases are capable of communicating and exchanging data (Sujanski 2001). Technically, a software component called a parser is required that analyses input data to build the underlying data structure which is often described as an abstract syntax tree or other hierarchical structure. The data structure enables different data and message formats (e.g. data-exchange protocols, programming languages) to be interconnected to an application programing 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.

However, consistent data and message formats do not ensure that the shared data are interpreted in a meaningful and accurate way by the ICT system. Here, semantic standards come into play. They usually harmonise and standardise terminology. A basic tool for doing this is a bio-ontology that formally represents knowledge in a given domain (such as cancer). The bio-ontology is concerned with which concepts are contained within the field, what information is required for each concept to exist, and how different concepts are related to each other. Therefore, bio-ontologies offer a structured knowledge repository that is used to describe the domain and can be used to draw conclusions about the entities within that domain.[6]

> *"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." (Interview 2, IT)*

Even if the interviewee prioritises the semantics, semantic and syntactic standards both are at work to facilitate data integration and to support the reliability and reproducibility of data and models. To reproduce data written in different formats and mapped from different bio-ontologies because of different source domains, metadata help to categorise 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. However, data integration not only requires information on how data are profiled, but also on how they will be used.

> *"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. How to do this is a very complex question ... 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." (Interview 9, BioMed)*

This citation implies that the use of data in research basically depends on chosen standards in data processing. Therefore, the standards that are at work define how the data may be used. However, due to technological innovations regarding ICT, multiple standard operating procedures (SOPs) are triggered at the same time (Auffray et al. 2009: 2). As a result, a growing number of overlaps and duplications of ICT-based standards evolve (Field et al. 2009: 234). Of course, anyone can claim to develop a new standard, but standards necessarily need approval by users. Therefore, the standards that are preferred are those that are supported by large user communities. It is assumed, for instance, that the ICT tools and services that will be widely recognized and reused are those that are built on broadly accepted standards.

———————

[6] 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).

> *"Well, the number of standards that you can use or that you could wish for is re-*
> *latively high. The problem is just that there are simply too 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 these areas, and in the end, everyone picks out whatever they happen to*
> *need. This means that 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." (Interview 13, IT)*

According to the interviewee, the ICT formats of existing tools work as standards for the creation of new tools. Hence, newly developed tools for scientific practice can be directly linked to the chosen ICT format. Taken the aspects outlined together, standardisation procedures in systems medicine are triggered by the need to make data and study designs comparable, ultimately in order to integrate and share data and study results. As data storing and processing is based on ICT systems, ICT also deploys sets of standards for data quality, annotation and exchange. Finally, ICT-based standards and guidelines define what counts as reliable evidence, clear nomenclature and commonly accepted experimental practices within the scientific community. Due to powerful standardisation procedures, the definition and use of data in scientific practice based on ICT is transformed. Data are split into at least two parts: the pure data content on the one side, and the metadata describing how the data were generated and how they are to be used on the other side. Of course, the distinction between data and metadata is a phenomenon with a long history in medicine and biology (Edwards et al. 2011; Leonelli 2010). What is different is the fact that the metadata in systems medicine are first and foremost defined and attached by ICT. It ultimately forms a new body of information including formal data and message formats, accurate classifications or other relevant ICT metadata. As a result, the significance and meaning of data have changed by defining which part is for scientific use and which one contains more or less purely technical information.

Conclusions

Integration as a prerequisite for mining and making use of heterogeneous data in systems medicine is often conceived as the major challenge in managing high-throughput data (e.g. O'Malley, Soyer 2012: 611). In our case study, the ACGT consortium tackled the challenge by setting up an ICT infrastructure addressing two levels of integration: syntactic integration provided the technological rack to facilitate data exchange, whereas semantic integration ensured that the data exchanged are accurately interpreted. The interplay of syntactic and semantic standards intends to enable reliable and valid comparison of entities and processes. In the case study, a range of technological tools and services has been developed to assist both dimensions of data integration. These technological tools and services are continually developed further on the basis of ICT standards, whereby existing tools work as standards for the new ones. Hence, ICT-based standards seem to be the central instrument of coping with the overarching tasks of data integration. Such data - produced by almost fully automated and highly standardised procedures - become

valuable primarily because of their reproducibility and reliability based on ICT (GarciaSancho 2012: 26). Life science data standardised by ICT can now be treated as a new, unified stock of data that can easily be accessed and distributed among the scientific community. Hence, data stored in ICT infrastructures that facilitate 'open access' among interested scientists are often seen as public resources, particularly when they are created and maintained through governmental funding (Leonelli, Ankeny 2012: 35).

Against this backdrop, it is not surprising that ICT infrastructures and digital databases have already gained a strong position among the systems medicine scientific community. They are often pictured as service facilities to ease research activities and collaborations in data and knowledge sharing within a research domain. At the same time, individual scientists and laboratories reject responsibility for those activities that are associated with data standardisation and integration as they are ascribed to the scope of the ICT infrastructures (Leonelli, Ankeny 2012: 34). From their perspective, ICT infrastructures are regarded as purely technical aids; only the actual use of the stored data is regarded to be a scientific endeavour.

The often-made separation of service and research as two independent endeavours, however, hides the fact that data management and its inbuilt requirements fundamentally influence mining, interpretation and conceptualisation of data. As ICT standards address syntactic as well as semantic integration, ICT sets 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. By assigning significance and meaning to data, the standards have an impact on how researchers think about classifications and experiments and therefore affect daily laboratory practice. By representing resources of knowledge and providing guidance for future action, ICT-based standards are not only powerful instruments to drive data integration and to facilitate research, but also systematically shape research practices from the outset. Thus, scientific practice in systems medicine has reached the mode of digitalisation.

Bibliography

Ankeny, R.A., Leonelli, S. 2011. What is so special about model organisms? Stud Hist Philos Sci: Part 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.

Bucur, A., Rüping, S., Sengstag, T., Sfakianakis, S., Tsiknakis, M., Wegener, D. 2011: The ACGT project in retrospect. Lessons learned and future outlook. Procedia Computer Science 4, 1119-1128.

Döring, M., Petersen, I., Brüninghaus, A., Kollek, R. forthcoming: Contextualizing systems biology. Presuppositions and implications of a new approach in biology. New York: Springer.

Edwards, P.N., Mayernik, M.S., Batcheller, A.L., Bowker, G.C., Borgman, C.L. 2011: Science friction: data, metadata, and collaboration. Soc Stud Sci 41(5), 667-690.

Fedoroff, H.J., Gostin, L.O. 2009: Evolving from reductionism to holism: is there a future for systems medicine? JAMA 302(9), 994-996.

Field D., Sansone S.A., Collis, A., Booth, T., Dukes, P., Gregurick, S.K. et al. 2009: Megascience. 'Omics' data sharing. Science 326(5950), 234-236.

Forgó, N., Kollek, R., Arning, M., Krügel, T., Petersen, I. 2010. Ethical and legal requirements for transnational genetic research. München: Beck.

Garcia-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.

Ge, H., Walhout, A.J., Vidal, M. 2003: Integrating 'omic' information: a bridge between genomics and systems biology. Trends Genet 19(10), 551-560.

Janes, K.A., Yaffe, M.B. 2006. Data-driven modelling of signal-transduction networks. Nat Rev Mol Cell Biol 7(11), 820-828.

Knorr-Cetina, K. 1999: Epistemic Cultures. How the Sciences Make Knowledge. Cambridge: Harvard University Press.

Lenzerini, M. 2002: Data integration: A theoretical perspective. Proceedings of the twentyfirst ACM SIGACT-SIGMOD-SIGART symposium on principles of database systems. New York: ACM Press, 233-246.

Leonelli, S. 2010: Packaging data for re-use: Databases in model organism biology. In P. Howlett, M.S. Morgan (eds.) How well do facts travel? The dissemination of reliable knowledge. Cambridge: Cambridge University Press, 325-348.

Leonelli, S. 2014: What Difference Does Quantity Make? On The Epistemology of Big Data in Biology. Big Data & Society 1. doi:10.1177/2053951714534395 (accessed March 4, 2015)

Leonelli, S., Ankeny, R.A. 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, A.D., Christie, K.R., Harris, M.A., Lomax, J. 2011: How the gene ontology evolves. BMC bioinformatics 12, 325.

Lynch, M. 1993: Scientific practice and ordinary action. Ethnomethodology and social studies of science. Cambridge: Cambridge University Press.

Meier S., Gehring, C. 2008: A guide to the integrated application of on-line 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 D. Garz, K. Kraimer (Hrsg.) Qualitativ-empirische Sozialforschung. Konzepte, Methoden, Analysen. Opladen: Westdeutscher Verlag, 441-471.

O'Malley, M.A., Soyer, O.S. 2012: The roles of integration in molecular systems biology. Stud Hist Philos Biol Biomed Sci 43(1), 58-68.

Philippi, S., Kohler, J. 2006: Addressing the problems with life-science databases for traditional uses and systems biology. Nat Rev Genet 7(6), 482-488.

Pickering, A. (ed). 1992: Science as Practice and Culture. Chicago: Chicago Press.

Rouse, J. 2002: How Scientific Practices Matter. Reclaiming Philosophical Naturalism. Chicago: University of Chicago Press.

Schmid, A., Blank, L.M. 2010: Systems biology: Hypothesis-driven omics integration. Nat Chem Biol 6(7), 485-487.

Sujanski, W. 2001: Heterogeneous Database Integration in Biomedicine. Journal of Biomedical Informatics 34(4), 285-298.

Swertz, M.A., Jansen, R.C. 2007: Beyond standardization: dynamic software infrastructures for systems biology. Nat Rev Genet 8(3), 235-243.

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.

Weston, A.D., Hood, L. 2004: Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine. J Proteome Res 3(2), 179-196.

Wierling, C., Herwig, R., Lehrach, H. 2007: Resources, standards and tools for systems biology. Brief Funct Genomic Proteomic 6(3), 240-251.

Yin, R.K. 2009: Case Study Research: Design and Methods. Thousand Oaks: SAGE Publications, Fourth Edition.