



Personalised Cancer Medicine Program at Karolinska Institutet

Report on the role of information technologies for the development of Personalised Cancer Medicine

Michele Masucci
23 February 2017

Personalised Cancer Medicine Program

Karolinska Institutet
Science for Life Laboratory
Tomtebodavägen 23A
171 65 Solna, Sweden

Office phone: +46-704-54 02 74
Email: info@pcm.ki.se
Webb: www.pcm-ki.se

Acknowledgements

Claes Karlsson
Ingemar Ernberg
Anders Wennborg

Ulla Wilking
Bengt Fundin
Shikako Suzuki
Luigi de Petris
Bertha Brodin
Loránd Kis

Olli Kallioniemi
Bengt Persson
Johan Rung
Valtteri Virta

Mariska Bierkens
Gerrit Meijer
Frank Ückert
Kristof von Kalle

Jeffrey Yachnin
Johanna Verneresson

Jesper Tegnér
Narsis Kiani

Anette Guy James
Bo Wikström
Björn Hultgrén
Staffan Lindblad

Dan Grandér
Roger Henriksson
Annelie Liljegren

Table of contents

- I. Summary
- II. Introduction
- III. Background
- IV. Procedures
- V. Selected Findings
- VI. Conclusions
- VII. Recommendations
- VIII. Appendices
- IX. References
- X. Bibliography
- XI. Glossary

I. Summary

The purpose of this report is to improve the understanding of the role of information technology in conducting translational cancer research with a precision medicine approach. The overall aim is to collect information and professional opinions on the current status of the clinical cancer research capability regarding information collection, management, analysis and dissemination.

Specific objectives have been to describe the current work situation regarding data handling for cancer researchers at Karolinska Institutet and Karolinska University Hospital conducting translational research projects with a precision medicine approach. A strong emphasis has been on the capability for conducting academic clinical trials and investigator driven research projects. Efforts have been aimed at identifying the primary working methods including popular software, hardware, data infrastructures, registries and databases as well as awareness on procedures for secure management and sharing of data. Precision medicine for cancer entails an intensification of the capability to generate, manage and analyse large quantities of heterogeneous sets of data to make it researchable and relevant for clinical decision-making.

The methods used to conduct the report has been mainly information gathering through literature review from scientific papers, interviews with members of the cancer research community in Stockholm as well as bench marking with international cancer centres regarding their information infrastructure for cancer research.

There have been several key findings as a result of this investigation. The main one being that there is a structural deficiency in the capacity to service advanced translational research mainly due to the lack of human resources with the right competence. The acknowledgement that shared infrastructures for clinical research is essential for improved results across research groups, profession and specific interests needs to be met with the appropriate investments.

The fragmentation within the cancer research community is reflected in the incapacity to join efforts in establishing collaborative platforms and resources that go beyond the specific interest of individual researchers. Among researchers, work on shared informatics has low priority. Time constraints limit the capacity to engage in work that would produce efficient infrastructures. Between clinical duties and research the capacity to designate focused work on informatics such as reaching consensus on terminology and content for clinical recordkeeping, designing standardised informatics workflows, or doing data quality assessments in collaboration with database technicians is limited. Although there are efforts from the hospital and KI to improve and facilitate clinical research - most notably the 4D for Breast cancer as well as the effort to standardise the clinical records in order to make reporting to quality registries more efficient – these efforts are designed as pilots that gradually are to be expanded, implemented and disseminated to all cancers. However, for the majority of the clinical researchers the concept of shared infrastructures and informatics support is practically identified as the hospital clinical trial unit, KPE, if not something that is solely an issue to organize by the individual researcher herself.

II. Introduction

Information technology can sometimes be conceived of as a magical solution to difficult problems. It is important to specify clearly what is intended with informatics, what kind of operation this technology is supposed to conduct and what results one is hoping to achieve. In cancer research, as in all medical research, a primary objective for productive hypothesis testing is to be able to collect and analyse high quality data on a relevant patient population. Throughout the history of modern medicine the capacity to provide evidence for a specific clinical measure has been based on the collection of data on patients.

IT-enabled health care can reduce fragmentation of information; ensure high-quality, safer care, and aggregate data to enable meaningful use at the point-of-care and population level. This can lead to a “national health IT ecosystem” in which every consumer, doctor, researcher and institution is important for improving care. This IT ecosystem would include electronic quality registries, a national cancer database, and electronic health record (EHR)–based tools for point-of-care decision support that can transform—in terms of pace, scale, and scope—the process of answering important cancer-related inquires. Its potential to do so lies in the inherent capacity to comprehensively capture rich patient data and to directly support care standardization. Technologic innovations from web-enabled mobile devices, integrated patient phenotype and genotype databases for individualised treatment, and real-time decision support, are gradually becoming introduced in clinical practice to enhance the clinical, organizational and relational aspects of care. Because many individuals are already within a care delivery practice or system at the time of cancer diagnosis, these tools can be useful across the continuum of cancer care to address prevention, diagnosis, treatment and survivorship.

This investigation is focused on collecting information surveying the conditions for translational cancer research from an informatics perspective. A number of leading questions have guided the investigation:

- “Which are the conditions for generating, collecting, managing and sharing research data of different kinds and from different sources within the cancer field?”
- “What IT systems are at place and what are their characteristics?”
- “What kinds of developments and future scenarios for the capacity to collect and curate high quality research data are there at the local and international context?”

A first phase in the investigation was to gather intelligence including survey of a number of leading cancer centers and their solutions for cancer research data. A model based on recent scientific publications on the topic of translational informatics was also a significant component¹. The second phase of the investigation covers interviews and a survey with stakeholder representatives and decision makers in the Stockholm region including the Karolinska Institutet, Karolinska University Hospital,

¹ Model was derived from “Translational Informatics Realizing the Promise of Knowledge-Driven Healthcare” edited by Payne, Philip R.O., Embi, Peter J. specifically looking at chapter 3 by Jessica T Tenenbaum on Personalized Medicine.

SciLifeLab, Stockholm Medical Biobank and other relevant stakeholders in the cancer research community. An important element of this has been to learn about the perceived needs from providers and the end users of research data and how this resonates with decision makers. Furthermore, the investigation has reviewed some technologies that could be implemented and resolve identified bottlenecks.

III. Background

Data, information, knowledge

Data is the lowest level of abstraction from which information and knowledge are derived. Information is the communication of data in an understandable way. Knowledge is derived from data and information through ‘formal or informal analysis’ – it is agreed to be true.

Information is an element capable of increasing or decreasing the certainty of a system. Data transforms into information by changing the level of certainty to a system. There are several dimensions to information that can be summarised into the technical, semantic and utility aspects. The technical aspects concern the capacity to store and share data in some form. The semantics regards the meaning; data needs to be intelligible in order to be informational. Information requires that the message is understood. Utility regards the fact that data needs to serve a purpose, to be a part in answering a question or contributing to an observation.

Computers are designed to process data and turn it into information. Technical advancements have provided increasing computational and operational power making it possible to analyse ever greater quantities of data with increased complexity and depth. The application of these advancements into biology and medicine has revolutionised the field during the past decades. Medicine becomes increasingly data intensive, demanding an IT readiness from the profession and the health care system. The capacity to transform data into information and then knowledge, the efficient processing of heterogeneous and complex data into evidence, has become more intensive. This has been shown not only by the increasing dependence on information systems in clinical practice, but perhaps most of all by the increasing use of molecular biomarkers requiring advanced bioinformatics when developed on the research side and when used clinically as diagnostic, prognostic or therapy predictive tools.

Legislation

The digitalization of research and medicine poses challenges to the protection of patient integrity and the safe management of sensitive data. There is an increasing public awareness of how collections of patient data can be hacked or misused. This requires that biomedical researchers and health care providers install safe systems and follow procedures for legally and ethically correct collection, management and dissemination of any data derived from patients.

There are a number of laws that govern the proper use of sensitive information on persons. The person data act is based on a EU directive². This law contains rules and regulations aiming to protect people from violation of personal integrity. In 2018 a

² European Commission Data Protection: <http://ec.europa.eu/justice/data-protection/> (2013-03-13)

new EU regulation on data protection will come into effect in all member states. This legislation aims at strengthening the rights for individual citizens by stricter demands on data security on data collectors and managers.

It is projected that these stricter rules and legislation demands that researchers and health care providers adapt their current praxis to better service the public with information of stored personal data and improved security measures.

The primary set of laws that govern translational research and clinical practice are:

- The Personal data act
- The Patient data act

Some of the major points from the Person data act are the following:

- People shall be protected against the violation of their personal integrity by processing of personal data.
- The registered person is entitled to information concerning processing of personal data that concerns him/her.
- A person who contravenes the Personal Data Act may be liable to pay damages or be sentenced to a criminal penalty.
- The processing of personal data shall be notified to the Data Inspection Board. However, this does not apply if the person who is responsible for the processing has appointed a personal data representative.

The Patient data act collects rules and regulation on information management in health care. The main principle for this law is that personal integrity should be respected within health care when care providers access information on the patient. This is secured by instituting selective access to information that is essential for the specific medical issue or department. Every health care provider has a responsibility to service care givers and patients with information systems that can provide the correct and secure handling of patient information.

In the transition from health care to research, legislation is generally transposed from an emphasis on the Patient data act to the Personal data act³ and the Data act. The real time generation and processing of information for clinical decision-making through electronic health record systems requires proper identification of the patient. At the moment of data generation for evidence collection of data in registries, clinical trials databases or other research databases primarily requires the correct association of diagnostic, treatment and outcome information of the patient, however not as an identified subject. De-identification is therefore a common practice. There are several methods for de-identification. Anonymisation is irreversible, meaning that identification with the source subject has been rendered impossible by erasing all possible data that could be used for identification or codes. De-identification can also

³ Personal Data Act (SFS 1998:204) from 24 October 1998. The Swedish government issued supplementary regulations in conjunction with the Act in the Personal Data Ordinance (1998:1191). The Data Act from 1973 is repealed by the Personal Data Act. Can be read in full here: <https://www.ics.uci.edu/~kobsa/privacy/Swedish.htm> (2017-03-13)

refer to data where the identity is possible to access in case it is needed but has been rendered immediately inaccessible to users. This kind of de-identification is considered weaker despite whatever method for protecting the possible linking data. However genetic data is considered specifically difficult to render de-identifiable as the actual genomic data is inherently a link to the person.

Data Security

Confidentiality and integrity of data are important; the availability of IT is also essential for scientific work. Therefore, redundant IT systems are needed in many places. IT has gained a pivotal role as an enabling technology in the life sciences and particularly in cancer research.

There is an increasing use of computerized systems in clinical trials to generate and maintain source data and source documentation on each clinical trial subject. Such electronic source data and source documentation must meet high standards of data security. The computerized systems should be designed: (1) to satisfy the processes assigned to these systems for use in the specific study protocol (e.g., record data in metric units, blind the study), and (2) to prevent errors in data creation, modification, maintenance, archiving, retrieval, or transmission (e.g., inadvertently unblinding a study).

There should be specific procedures and controls in place when using computerized systems to create, modify, maintain, or transmit electronic records, including when collecting source data at clinical trial sites. In translational research projects there are some aspects of data security that are essential.

Data access

Access must be limited to authorized individuals. External safeguards should be put in place to ensure that access to the computerized system and to the data is restricted to authorized personnel. Staff should be kept thoroughly aware of system security measures and the importance of limiting access to authorized personnel.

Time stamps

Computer-generated, time-stamped audit trails or other security measures can capture information related to the creation, modification, or deletion of electronic records and may be useful to ensure compliance with the appropriate regulation.

Controls should be established to ensure that the system's date and time are correct. The ability to change the date or time should be limited to authorized personnel, and such personnel should be notified if a system date or time discrepancy is detected.

System Documentation

For each study, documentation should identify what software and hardware will be used to create, modify, maintain, archive, retrieve, or transmit clinical data.

System Controls

When electronic formats are the only ones used to create and preserve electronic records, sufficient backup and recovery procedures should be designed to protect against data loss. Records should regularly be backed up in a procedure that would prevent a catastrophic loss and ensure the quality and integrity of the data. Records should be stored at a secure location. It is important to maintain backup and recovery

logs to facilitate an assessment of the nature and scope of data loss resulting from a system failure.

The problem of identification with genomic data

The one area in which genetic tests are qualitatively different is that even a small fraction of an individual's whole genome is highly identifying. This makes the stakes of data security and privacy policies much higher. Genomic data cannot be shared with other researchers without a much higher likelihood of disclosure of identity than with other data types. The risk of disclosure is growing significantly with the accumulation of independent DNA databases.

Training of Personnel

Those who use computerized systems must determine that individuals (e.g., employees, contractors) who develop, maintain, or use computerized systems have the education, training and experience necessary to perform their assigned tasks.

Training should be provided to individuals in the specific operations with regard to computerized systems that they are to perform. Training should be conducted by qualified individuals on a continuing basis, as needed, to ensure familiarity with the computerized system and with any changes to the system during the course of the study.

Conditions for conducting PCM research

Conducting research with a Personalised Cancer Medicine concept is a complex and time-consuming effort. There are several challenges to the PCM endeavour: patient recruitment, design of the trial, administrative and legal standards need to be met, and funding needs to cover analysis and personnel costs. Since these efforts amount to a complex web of decisions and information flows, projects are often set up to answer the specific research question according to the unique constellation of researchers. However, a set of needs and issues are often recurring based on the local infrastructure

PCM research faces two important challenges. The first is organisational: bringing together researchers from many countries, achieving consensus and overcoming the many regulatory and financial barriers which can impede the smooth running of international clinical research. The second is methodological: even with international collaboration, standard trial designs may require unfeasibly large recruitment targets for the setting, which calls for innovative methodologies to be researched.

Information technology for -omics

The introduction of Next Generation Sequencing in routine health care is already reality in some progressive cancer centres around the world. Genomic test data are very complex. However, this does not constitute a unique situation in medicine, there are other complex tests that have been implemented such as advanced Imaging.

Genomic tests, like all other clinical tests, provide a probabilistic measure of certainty that a specific pathophysiological state is present (i.e. diagnosis) or will be present (i.e., a prognosis). Whether genomic or conventional, all these tests are used for clinical decision making whether in the context of screening asymptomatic individuals or managing individuals with a complaint. The costs of each of the

aforementioned tests, including whole genome sequencing, are typically less than a couple of thousand dollars and the volume of data generated is typically not more than a few terabytes.

FAIR data principles

The independent research community FORCE11⁴ is constituted of scholars, librarians, archivists, publishers and research funders that want to facilitate the change toward improved knowledge creation and sharing. One of the grand challenges of data-intensive science is to facilitate knowledge discovery by assisting humans and machines in their discovery of, access to, integration and analysis of, task-appropriate scientific data and their associated algorithms and workflows. Their proposal is that researchers and data managers design and develop systems according to a set of principles summarized as the FAIR data principles.

To be Findable:

- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
- F2. data are described with rich metadata.
- F3. (meta)data are registered or indexed in a searchable resource.
- F4. metadata specify the data identifier.

To be Accessible:

- A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
 - A1.1 the protocol is open, free, and universally implementable.
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
- A2 metadata are accessible, even when the data are no longer available.

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles.
- I3. (meta)data include qualified references to other (meta)data.

To be Re-usable:

- R1. meta(data) have a plurality of accurate and relevant attributes.
 - R1.1. (meta)data are released with a clear and accessible data usage license.
 - R1.2. (meta)data are associated with their provenance.
 - R1.3. (meta)data meet domain-relevant community standards.

Despite the fact that the FAIR data principle is far from a global standard, this list of objectives are an important guide in the formation of multicenter collaborations and the data design of research projects. The FAIR principles help in both the in house set-up and use of data and for facilitated data sharing.

IV. Procedure

⁴ A community of scholars for improved information and knowledge creation.: <https://www.force11.org> (2017-03-13)

The report has been conducted by the collection of information through a variety of means.

Literature review

Extensive reading of current literature on translational informatics for cancer research was conducted to ground the investigation in line with recent developments in the field.

Observational studies

Practice based information gathering through the involvement in translational research projects, design and management of research projects involving issues concerning data gathering and management.

Interviews

Interviews with main stakeholders such as cancer care providers, translational cancer researchers, basic researchers, health care management and research management, legal experts, and IT experts involved in the development and management of cancer research infrastructures.

Workshops and site visits

Physical and online meetings with representatives of IT departments from Deutsches Krebsforschungszentrum (DKFZ)⁵ and Netherlands Cancer Institute (NKI)⁶ regarding issues on data collection, analysis, storage and management.

V. Selected Findings

A summary of the main findings that are presented below. These findings constitute the foundation for the conclusion and subsequent recommendations of the report.

The DKFZ Heidelberg Case

The DKFZ information technology (IT) needs and requirements are continuously growing. This is due, in particular, to modern laboratory technologies in areas such as genome analysis and radiological image processing, which generate huge amounts of data on a petabyte scale. Furthermore, as more and more personal data are being processed, aspects of IT safety play an ever more important role in many departments of the DKFZ and as a requirement in many national and international science projects and collaborations such as the German Consortium for Translational Cancer Research (DKTK).

The task of the Information Technology Core Facility (ITCF) is to optimize the use of IT at the DKFZ, to advise all in-house departments in selecting IT tools for their specific needs in a general context, to implement and operate centralized IT services as well as to support the planning and operation of individual solutions. This is done on a service-oriented basis. All essential strategic and operative functions and tasks at the DKFZ are substantially supported by IT.

⁵ DKFZ IT Core Facility: <http://www.dkfz.de/en/itcf/index.html> (2017-03-13)

⁶ NKI Research IT: <https://www.nki.nl/research-it/research-it/> (2017-03-13)

The Software Systems working group provides software for server systems and applications for client systems. This includes central services like user management, software distribution, license management, access- and resource-management and administration for the central database server. The “DKTK” team is responsible for DKTK Users, DKTK intranet and DKTK projects.

Data acquisition is guided by the principle that data should remain in local data repositories under the governance of data owners and made accessible according to established conditions. This idea for data warehouse solutions is built on networked data repositories and federated databases.

The German Federal Government provides 90 percent of the funds for DKTK, and the State of Baden-Wuerttemberg provides the remaining 10 percent. Further financial sources are external funding, license revenues, and donations.

The Netherlands Cancer Institute NKI Research IT Facility

The Research IT facility has the mission to develop solid and sustainable Information Technology (IT) infrastructure to provide state-of-the-art IT services to NKI researchers.

Services that are offered:

- High Performance Computing (HPC) and Storage on local HPC servers
- A source code repository for researchers and projects
- Access to the national Life Science Grid of SURFsara
- Hosting of research group and publication websites
- Archiving of research data sets
- Data integration and project management consultancy (focusing on clinical and molecular data)

In development are:

- A patient centric data warehouse aggregating (and integrating) clinical and research data sources
- Front end for researchers to inspect the data catalogue and make data requests
- A data stewardship platform for securing and sharing research data
- Front end for Principle Investigators, researchers and data managers to manage data for projects, research groups and core facilities

Goals are achieved by participating in local, national, and international projects. The team adopts the SCRUM⁷ methodology. This is a software development method that works iteratively with a flexible and holistic product development approach.

The IT services of the NKI are a core service and covered by institute budget which has a basic funding from the Dutch government together with project grants and funding from the KWF Dutch Cancer Society.

The Karolinska Clinical Trial Office Karolinska

The Oncology Clinic has a clinical trials office that assists in organizing clinical trials

⁷ SCRUM Alliance: <https://www.scrumalliance.org/why-scrum> (2017-03-13)

management for academic studies and Industry sponsored clinical trials. Each project has its unique conditions, however there are some standard procedures regarding patient inclusion to trials and data collection that have been established. The organization employs research nurses, statisticians and data managers that assists in the design of electronic Case Report Forms (eCRF) through the use of licensed software. Data collection, quality assessments, monitoring coordination and data storage and sharing follow common procedures for clinical trials. There is no specific readiness or preparation concerning adaptation of more innovative IT solutions other than the manual collection of data and input through eCRF:s into databases.

4D theme Breast Cancer

The Stockholm County Council is investing in informatics development with the ambition to achieve coordination of heterogeneous data sources that are used through the care pathway for individual patients. In the projects 4D referring to the first four diagnoses that were chosen as pilots one of the main ambitions is to improve data quality and access. The aspirations are improved decision making and early detection, improved patient information and better access to data for research purposes and outcomes. Having an overview of where data is generated and stored and how this data can be accessed is essential for both improved care and the capacity to answer research questions. The breast cancer care pathway is the first pilot among the different cancer. The results and methodologies that come out of the pilot will be reproduced for other patient groups.

cBioPortal

Through the Cancer Core Europe initiative, a number of research softwares and data sharing solutions have been discussed. An example is the open source tool for exploration of cancer genomics data sets cBioPortal⁸. The cBioPortal for Cancer Genomics was originally developed at Memorial Sloan Kettering Cancer Center (MSK).

The cBioPortal for Cancer Genomics is an open-access, open-source resource for interactive exploration of multidimensional cancer genomics data sets. The cBioPortal significantly lowers the barriers between complex genomic data and cancer researchers who want rapid, intuitive, and high-quality access to molecular profiles and clinical attributes from large-scale cancer genomics projects and empowers researchers to translate these rich data sets into biologic insights and clinical applications.

VI. Conclusion

General role of IT

The need for improved IT know how is clear. However, the needs are diversified depending on the phase of the translational research project. Some research projects need bioinformaticians, others infrastructure support, others need general information on standards and methodology. Data acquisition for clinical trials follows certain well documented routines and protocols often negotiated and set up in collaboration between hospital management, investigator and industry sponsors. Implementation

⁸ cBioPortal, developed by Memorial Sloan Kettering NY: <http://www.cbioportal.org> (2017-03-13)

studies and outcomes research studies depend on the capacity for the healthcare system to provide high quality data on diagnosis, treatment and outcomes. Furthermore, research project trying to answer specific question based on bio samples and related clinical data are often structured at an ad hoc manner depending on the specific project, people involved, funding and access to data and samples.

Implementation of standards

The software solutions used are constantly developing; this makes it difficult to define a field standard. Different software produces different output, making it difficult to assess and choose among service providers. This poses a challenge for transnational collaborations as well as collaborations on regional and even local levels. The question of standards, and standardisation, together with the costs and efforts related to this challenge are recurring.

Structured Care Data Initiative

There are issues concerning IT that are common to all forms of hypothesis driven investigation, whether it is related to testing new treatments on patients, conducting lab work on biological samples or evaluating the efficacy and efficiency of cancer care. Certain clinical attributes from the patients are often reoccurring in answering research questions. These are central information items that are needed to evaluate cause and effect over time. Diagnostic information and health status, together with disease progression and technical information concerning the treatment are essential. It is therefore important that these attributes are well documented and that the information systems used for their registration facilitate input with for example data verification scripts. The quality of data is always best at first entry; secondary interpretation of unstructured data always opens for misinterpretation and is generally time consuming.

There are important efforts being made to achieve standardisation of the record keeping at the oncology clinic Radiumhemmet with the help of IT specialists at the Karolinska Hospital in collaboration with practicing oncologists. The aim is to improve record keeping and reach a structure that can make automated data extraction possible. Data localization, extraction and further load into new data repositories require that data follow predictable formats that can be identified, located and transferred. TakeCare is the electronic record keeping system used by Karolinska University Hospital and the New Karolinska Hospital together other health care providers. There are modules and functionalities that facilitate structured data in this system. However, it requires informed development by care providers that can reach a consensus on standard formats for chosen attributes. Data that is generated through affiliated systems such as radiology assessments and pathology reports enter the electronic records in raw format and are often commented in free text. The need to standardise these reports could make the interoperability and data transfer more secure both for readability and further data extraction.

Manpower and competence

There is a general lack of personnel with IT competence that can assist in infrastructure development and support within the cancer research community. Researchers specialised in system development, programming, IT system management and informatics are needed. IT services available for clinical researchers

through the hospital and clinical trial unit are sufficient for the needs of on going clinical trials.

There is a low level of innovation aiming at improving patient recruitment data quality assessment and automatic data retrieval for clinical studies.

Research and development funding regarding this type of infrastructure is focused on the pilot project 4D Breast cancer⁹ which are aiming at establishing an integrated information infrastructure for clinical cancer care in breast cancer that also can serve research.

There is an obvious lack of infrastructural support related to current electronic patient medical records for the acquisition of and processing of genomic and genealogical data. The best efforts are focused on structuring through the Structured Care Data¹⁰ project, however this covers only a selection of cancers and the compliance among oncologists is not complete. Individual developers disseminating information and developing the tools for structured record keeping have made heroic efforts. However the new procedures can still be difficult to adapt to for some clinicians. Full compliance at the local oncology clinic and across cancer clinics in the region will require more development, assistance and educational efforts.

VII. Recommendations

Recommended PCM Program Actions

There are several concrete actions that the PCM Program can take to improve the development of IT for PCM research and care. The PCM program does not formally own research data nor any form of research infrastructure by itself. However, the PCM program has the potential for leveraging an optimal development through its role of associating clinical researchers and the cancer research community together with decision makers and international collaborators.

1. Continuous information collection of the IT needs such as quality of data, software and hardware, for on-going and prospective clinical trails and research projects.
2. Continue to survey ongoing developments in data collection and infrastructures for clinical cancer research.
3. Facilitate dissemination of information across stakeholders on new standards and methods for data collection and management in biomedical research focusing on cancer and precision medicine.
4. Support key stakeholders in implementation of new technologies and methods for data input, extraction and safe sharing. This includes assisting in coordination of different initiatives at a local, national and international level.

⁹ Program 4D: <http://www.sll.se/verksamhet/halsa-och-varld/Aktuella-projekt-Halsa-och-varld/Program-4D/> (2017-03-13)

¹⁰ Documentation on Strukturerad Vårddata:

http://www.termkonferens.se/2011/Termkonferens_2011_CSTC_Take_Care.pdf (2017-03-13)

5. Establish a local IT Task Force consisting of a group of professionals with relevant competence and background. This task force will engage in ongoing IT related issues and respond to requests and tasks from local research projects, Cancer Core Europe and other collaborative entities.

Support Cancer Core Europe Activities

A continuous participation in Cancer Core Europe activities is essential to boost the development of shared infrastructures and practices for translational cancer research projects at Karolinska. This includes further development on the following ongoing efforts:

- Research discovery project on patients with somatic BRCA1/BRCA2 mutations in non-breast tumor types by sharing patient data among Cancer Core Europe centers
- Surveying the IT infrastructure capability at each CCE member site in order to develop data sharing procedures.
- Participate in applications for funding of IT infrastructures for data sharing across CCE centers as a model for future multinational cancer research in Europe.
- Assist with IT issues for the Basket of Basket Clinical Trial that connects all CCE members.

Developing a meta data repository

Computing infrastructure can be designed for scalability and flexibility utilizing the latest information technologies, to enable us to meet the rapidly increasing storage and performance requirements of data driven research. Integration of patient cohorts and related omics data using dedicated research infrastructures to deliver consistently faster, excellent and unique outputs when utilising layered data for stratified and personalised medicine enhances uptake, cost-effectiveness and resilience and will facilitate the use of expanded resources and knowledge. However, competition in funding and prestige creates a general scepticism for participating in the development and use of shared infrastructures for the collection of sensitive unpublished research data. A possibility to overcome this resistance is to clarify the benefits and gains such as cost reduction and access to competence and state of the art data storage systems for heterogeneous data sets. The PCM program could also service clinical cancer researchers by providing a meta data repository collecting contextual information on research data sets and clinical trials. This could provide continuity and coordination and form a first step in more concerted actions on a local level strengthening the capacity for successful completion of research projects. This effort would require investment in hardware and software together with the development and management competencies.

VIII. Appendices

1. Questionnaire CCE IT Data Sharing Task Force
2. CCE Pilot Project description
3. Proposal for data integration and sharing tool with Johan Rung for SciLifeLab

IX. References

Leslie G. Biesecker, et al. *Next Generation Sequencing in the Clinic: Are we Ready?*
Nat Rev Genet. 2012 Nov; 13(11): 818–824.

Payne, Philip R.O., Embi, Peter J. *Translational Informatics Realizing the Promise of Knowledge-Driven Healthcare*

X. Bibliography

- Guidance for Industry - Computerized Systems Used in Clinical Investigations (FDA) Office of the Commissioner (OC) May 2007
- Cancer Research at DKFZ 2016
- Information on the Personal Data Act – The Swedish Government
- Patientdatalagen och den personliga integriteten

XI. Glossary

Electronic Record: An *electronic record* is any combination of text, graphics, data, audio, pictorial, or other information representation in digital form that is created, modified, maintained, archived, retrieved, or distributed by a computer system.

Original data: For the purpose of this guidance, *original data* are those values that represent the first recording of study data.

Direct Entry: *Direct entry* is recording data where an electronic record is the original means of capturing the data. Examples are the keying by an individual of original observations into a system, or automatic recording by the system of the output of a balance that measures subject's body weight.

Computerized System: A *computerized system* includes computer hardware, software, and associated documents (e.g., user manual) that create, modify, maintain, archive, retrieve, or transmit in digital form information related to the conduct of a clinical trial.

Source Documents: Original documents and records including, but not limited to, hospital records, clinical and office charts, laboratory notes, memoranda, subjects' diaries or evaluation checklists, pharmacy dispensing records, recorded data from automated instruments, copies or transcriptions certified after verification as being accurate and complete.