

Standardizing the Transmission of Genomic-Related Medical Data for the Optimization of Diagnostics and Research in the Field of Oncological Diseases

Elisabeth Pantazoglou[#], Martin Thoma[#], Lasse van de Sand, Heike Dewenter, Sylvia Thun

Niederrhein University of Applied Sciences, Faculty of Health Care, Competence Center e-Health, Krefeld, Germany

Abstract

Introduction: In Germany, there is currently no consistent analytic structure within genomic diagnostics in oncological diseases. Within the framework of the project GENeALYSE, a standardized and interoperable specification for associated uses cases shall be developed.

Intended Methods: Through process analysis and interface modeling, problems of the actual processes will be depicted between the involved actors. In the next step, the workflows and relevant findings will be displayed and adapted. In particular, the heterogeneous workflows in genome diagnostics will be represented by semantic annotation in an international terminology. The results of the semantic annotation build the basement for the creation of an implementation guide for standardized genome analytics, referring to HL7 Clinical Document Architecture (HL7 CDA).

Discussion: The problems of heterogeneous genomic diagnostics as well as unstructured findings in oncology leave the actors face comparable challenges on a regional

and supranational level. Interfaces, ambiguous semantics and manual activities inhibit interoperability, promote errors and lead to risks for patients and their sufficient medical treatment. A major challenge will be consistency between the heterogeneous terms to be found in genome analysis. The problem shall be addressed via using international terminologies as well as appropriate mapping techniques.

Conclusions: The aim of the project is to create an implementation guide for standardized digital documentation and communication solutions between diagnostics, medical therapy and research in the field of genome analysis. GENeALYSE is intended to optimize the coordination between the diagnostic genome laboratory and the clinical therapy decision in order to increase the safety and success of medical treatment, as well as to improve the health-related quality of life of the affected patients.

Keywords

Interoperability; Semantics; Clinical Document Architecture (CDA); Genetics

Correspondence to:

Elisabeth Pantazoglou, B. Sc. Health Care Management
Niederrhein, University of Applied Sciences, Faculty of Health Care,
Reinarzstrasse 49, Krefeld, Germany.
E-mail: elisabeth.pantazoglou@hs-niederrhein.de

EJBI 2017; 13(1):67-71
received: June 09, 2017
accepted: July 16, 2017
published: October 10, 2017

[#] Contributed equally

1 Introduction

Currently, cancer is one of the most common diseases in the German society. Based on statistic data from the National Statistical Office in 2013 breast cancer, lung cancer and colorectal cancer were identified as the most frequent causes of death for women placed on rang six, seven and ten. For men the most frequent causes of death were lung cancer, colorectal cancer and prostate cancer on rang two, six and seven [1]. In the same year the incidence of cancer reached 482.000 new cases and a total of 223.000 patients died because of this disease. The ten-year prevalence was

2.6 million. According to a report about cancer in Germany published in 2016, about four million people suffered from cancer in their life time [2].

The underlying causes of cancer are complex, but the main reasons for the rising number of newly reported cases of cancer are a rapidly increasing life expectancy and the reduction of other life-threatening diseases. In general, there is also a discussion about consumption of alcohol and tobacco as well as certain aspects of nutrition relating to cancer incidences [3, 4].

Nowadays the most common therapy against cancer is chemotherapy. With this treatment approach chemical

substances, the so-called cytostatic drugs, are administered to the patient with the aim to stop the division of the malignant cells. This form of therapy can be regarded as extremely radical, because cytostatic drugs do not specifically affect cancer cells, but also every form of cells inside the human body. In this context chemotherapy can be described as a severely impairing therapy option [5].

Even if chemotherapy, especially in association with radiotherapy, is very efficient on cell destruction there are very strong side effects for the patients and a high risk of an absent treatment success or the return of recurrences. The patients suffer extremely with this treatment. Because of the massive effects and adverse side effects of chemotherapy the patient's quality of life drops rapidly. In some cases these effects cause other disorders or severe health consequences for the patients. However, the lack of a therapeutic effect is the most serious problem in cancer treatment. The reasons for reduced treatment success are often resistant cancer cells or a late diagnosis of the cancer disease [6].

Advances in molecular biology have brought new information about the development of cancer in recent years. The genetic disposition of a person gets an important role along with external risk factors in the incidence of cancer diseases. In addition, the molecular-genetic diagnostics of tumor cells provides accurate information about possible entry points for efficient therapy options. This has led to new methods in cancer treatment, the so-called "personalized medicine". The new methods of genome sequencing are summarized by the concepts "Next Generation Sequencing"(NGS) and "Precision Medicine". Combined with medical databases they possess the potential of predicting the probability of cancerous diseases and to apply more specific therapy forms and therefore they may improve the quality of life and increase the survival rate of the affected patients [3, 7].

The new methods of the upcoming medicine become more predictive and preventive, as they contribute to a new medicine concept. They form the basis of the paradigm shift in modern medicine far away from the reactive medicine towards a more proactive and personalized medicine, the so-called P4-Medicine [8]. Instead of a general treatment the personalized medicine focuses on two different approaches. The first is used to detect the patient's potential risk of developing cancer and to initiate preventive steps. For this genome sequencing blood-material or other physical material is needed. The second is used for genome sequence analyses to verify mutations and to identify possible attack points. For that genome sequencing tumor material is needed. In addition to a specific requirement for the detection of the existence of a certain mutation in association with an individual clinical implication, it is possible to use whole genomic (sub-) sequences for diagnostics. This contains the risk, that the requirements to the molecular genetic laboratory are not clearly expressed. Furthermore the interpretation of the determined genome sequence depends especially on the physician's or natural scientist's knowledge.

However, funded diagnostics require acknowledged reference databases, which may be license- or charge bound. Though the databases support the medical findings, the problem remains, that they for free available platforms a standard procedure doesn't exist [9]. Another difficulty lies in the different output formats and the use of different annotations of genome sequences in the existing databases. A further problem lies in the report of the diagnostic test results. Neither the generation nor the transmission of the report is standardized at the current time. Solely narrative text modules can be used to support generating the diagnosis report. Along with the lack of syntactic standards, semantic reference systems are not applied regularly. Furthermore, the transmission of molecular genetic diagnosis reports is still paper bound [10]. Due to a non-consistent term definition these circumstances may affect the therapeutic decisions for the patients.

The aim of the future research project GENeALYSE is to create an implementation guide for standardized digital documentation and communication solutions between diagnostics, medical therapy and research in the field of genome analysis. GENeALYSE is intended to optimize the coordination between the diagnostic genome laboratory and the clinical therapy decision in order to increase the safety and success of the treatment, as well as to improve the health-related quality of life of the affected patients. The project will be coordinated via the Niederrhein University of Applied Sciences Krefeld, in collaboration with the Institute for Pathology at the University of Cologne and the Clinic for Gynecology and Obstetrics at the University Hospital of Düsseldorf. The located Center for Familial Breast and Ovarian Cancer (ZFBK-D) is part of cooperation between 17 university centers and is specialized in genomic diagnostic for genetic induced family-related cancer. The Institute for Pathology at the University of Cologne diagnoses genetic modifications on solid tumors. In addition it is a central molecular diagnostic center for patients with lung cancer in the network "Genomic Medicine" in Germany. This collaboration ensures that the findings of genome sequences consider all medicinal domains. The project is also supported by the involvement of associated partners, such as the Cancer Society North Rhine Westphalia (NRW), the Scientific Institute of Office-based Hematologists and Oncologists (WINHO) and the Federal association of German pathologists. The project is scheduled to start in September 2017 with duration of two years and it is funded by the Ministry of Innovation, Science and Research from the federal state of North Rhine – Westphalia (NRW).

2 Intended Methods

2.1 Process Analysis

The methodological basement will be the development of technical basics and the collection of information about

the actual processes. In the beginning, legal and regulatory framework conditions will be analyzed, which significantly influence the medical findings in genome analysis. Within the collection and analysis of the actual processes, scientific methods, e.g. analysis of the vulnerability, stakeholder analysis and questionnaire methodology, the specific workflows and the participatory environment will be evaluated. The goal is to examine the relevant workflows and to represent the main actors and stakeholders.

Associated information will be collected and illustrated using the Business Process Modeling Notation (BPMN) as a diagram Table 1.

Together with the project partners as well as selected experts from the specific areas, the listed questions will be discussed within several project meetings. The areas of clinical requirements, data transmission, sample entry, assessment of findings will be discussed as the main content inside the project meetings. Particular attention is also paid to the requirements concerning the tumor tissue. In contrast to the germ line determination of biomarkers, tumor tissue cannot be easily retrieved. Against this backdrop it is necessary to precisely define requirements for removal, transport and medical questionings concerning tumor tissue. The search for alternative possibilities for action or rather deviating processes as well as the collection and evaluation of exemplary results and reports of results from the experts involved and other relevant materials complete the survey of all necessary information and form the basis of the weakness analysis. Furthermore, the technical conditions will be analyzed. This includes the recording of the involved information systems, the necessary interfaces and transport

standards for data transmission. This methodological approach forms the basis for the further elaborations and involves the perspective and the requirements of the actors involved as well as the legal framework. Based on the actual process recording, optimization potentials will be derived and documented in the process of genome analysis. This will determine the requirements for the following target process determination and conception.

2.2 Process Modeling

Using the collected and analyzed information about the actual processes, potentials will be derived and presented via BPMN. The identification of the target processes have major influence on the development of the semantic reference systems and the implementation guide. In addition, requirements and framework conditions for necessary software and hardware must be formulated. The quality management approach “Six Sigma” with its core concept “Define-Measure-Analysis-Improve-Control” (DMAIC) and the associated tools will be used to create the target processes by following and combining the scientific methodology from the previous steps. In several individual project meetings expert opinions will be incorporated into the development of the target processes [11, 12].

In comparable mode of operation, the research topic will be limited, then separately considered and formulated, and finally validated in the review process. The resulting requirements for the target processes form the basis for the semantic annotation, as well as the creation of an implementation guide. For the modeling, a continuous exchange with the involved actors is indispensable. The work will be reviewed and adapted by a continuous process of improvement in accordance with KAIZEN to ensure the best possible applicability for all stakeholders. Documentation-based quality assurance is an essential part of research and development projects, in order to understand, analyze and correct mistakes in individual steps and changes [12].

2.3 Semantic Annotation and Implementation Guide

Against the backdrop of the technological developments in the field of NGS and related comprehensive genetic diagnostics, the interpretation of the identified sequence variants relating to their possible pathogenicity is very challenging. A current, structured and standardized data query of the existing knowledge for a functioning classification of DNA sequence variants provides the foundation for evidence-based clinical treatment options and therapy decisions. Especially the molecular diagnostic for the detection of a genetic predisposition requires an evidence-based interpretation of the analysis results. Furthermore only the standardization of the diagnosis report with semantic clear terms as well as a standardized communication of the

Table 1: BPMN-displayed questions.

Clinical questions and requirements Which information is needed by the requesting physician? Which information is available about the patient? How is the data collected and recorded?
Transmission of requirement and sample material How is the transmission performed? Do certain times or conditions of delivery have to be observed?
Delivery of the samples in the laboratory Which tests are carried out? Which questions or problems arise often?
Clinical results Which methods are implemented? Which support systems are used?
Assessment and reporting How is the result obtained? How is the report submitted? Which semantic and syntactic standards are used? What are the contextual problems or questions?

reports may avoid misinterpretation like wrong therapeutic or preventive decisions. The technical complexity of this medicinal field and the substantial amount of the collected information is a strong challenge for standardization. However, due to the rising relevance of these diagnostic methods and new possibilities in this field, standardization is indispensable.

Semantically clear terms generate the basis for a uniform, content-based understanding in the area of genome analysis. This is the basic prerequisite for the successful cooperation of all stakeholders. Semantic interoperability ensures a smooth exchange of information with a lower risk of misinterpretation. Therefore an analysis of available semantic standards for the representation of gene analyzes is planned. The survey is carried out via scientific literature analysis and in exchange with the experts. This forms the prerequisite for the development of the matching characteristics between the terminologies and their mapping.

First researches have shown that the HGSV nomenclature (Human Genome Variation Society) is a worldwide-adopted standard for naming genetic sequences. It is the result of the so-called "Sequence Variant Description Working Group" (SVD-GV), which is composed of the specialist groups Human Genome Variation Society (HGSV), Human Variome Project (HVP) and the Human Genome Organization (HUGO). The HGSV nomenclature is used for the uniform annotation of specialist terms in genome sequencing. The standard describes among other things the designation of the determined sequence in a particular format as well as in connection with a reference sequence [13].

The use of the nomenclature is also becoming increasingly important for research in databases and is increasingly being demanded in the field of communication [14]. Logical Observations Names and Codes (LOINC) is a nomenclature for laboratory analyzes and clinical observations, which has been developed by the Regenstrief Institute since 1994 and is available free of charge via a database. LOINC terms follow also the recommendations of other nomenclatures such as HGSV. The use of the LOINC Codes is provided in HL7 Clinical Document Architecture (CDA) documents. The ISO-Standard CDA is used to exchange electronic clinical documents in a standardized form. These two standards provide the framework for structured reporting of laboratory test results. In conjunction with gene sequencing they can be used for reports to structure semantically interoperable findings. For this purpose LOINC and HL7 collaborated in two HL7 Clinical Genomic Working Groups to define implementation guides for structured reporting of genetic tests [15]. These implementation guides shall be adapted for German conditions.

Terminological experts carry out the medical coding and the creation of the mapping tables independently. In addition, the ART-DECOR software serves as an online tool for the

experts' quality assurance. ART-DECOR is a web-based open source tool which supports the creation and maintenance of CDA templates, value sets, codes and data sets. Its underlying data format is XML. With an extension it can be used for the quality assurance. In addition, the semantic coverage (Figure 1) of the specialized terms as well as the quality assurance will be carried out by classification according to ISO TR 12300 [16]. The mapping tables will be used during the diagnostic evaluation and make a significant contribution to facilitating the currently highly complex searches in scientific genomic databases. Furthermore, the semantic annotation is part of the CDA (Figure 2), which will be represented in ART-DECOR.

The harmonization with international standards such as HL7, as well as e.g. LOINC, HGSV or the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT) forms the basis for the creation of the implementation guide. The work will be continually presented at the German interoperability forum and coordinated by the involved experts. Finally, the implementation guide shall provide all necessary information regarding structure and semantics for the standardized and interoperable exchange of results in the genome-sequencing field.

3 Discussion

Although genetic findings in oncology have for some time been a basis for the therapy decision, these results are still unstructured and differ from laboratory to laboratory. Against this backdrop, the lack of structure generates a huge problem in the interpretation of results as well as in research for therapeutic approaches. The unstructured

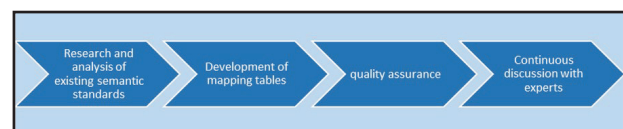


Figure 1: Process of elaboration of the semantic coverage.

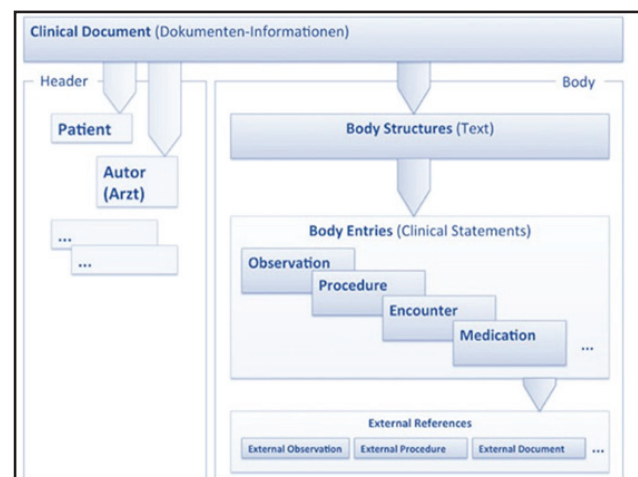


Figure 2: Simplified overview of a part of the CDA model.

report of genomic analyzes in oncology is not a regional problem. Nationwide, the actors face comparable challenges, leading to countless semiprofessional approaches and therapy solutions. The progress and the gain in knowledge in the fields of oncological laboratory analysis as well as the standardization of medical terminologies are progressing rapidly. The current findings are not standardized and developed in different forms depending on the institution. In addition, the results are provided in different quality, depending on the performing laboratory. If unstructured or incorrect information is passed on in the assessment, this may have negative effects on the treatment and therefore on the quality of life and ultimately on the survival of the patients.

With standardized digital techniques, genetic information about cancer cells as well as on biomarkers can be transferred and evaluated in high data quality and with little range of interpretation. Clearly and unambiguously structured results of the genomic sequencing are the basis for the search for targeted, individual therapeutic approaches for each patient. Thus the finding is the most important starting point and the decisive basis for the ultimately to be used therapy. In this project the findings from the fields of analytical laboratory diagnostics with the focus on oncogenes, the scientific analyzes and the development of innovative technical solutions in the eHealth sector as well as the standardization of communication in the health and social sector will be brought together. These three areas can be the basis for improved data quality and contribute to the improvement of communication by establishing a uniform findings structure.

In this way, the treatment quality will be positively influenced. The development of a structured electronic finding on the basis of semantic uniqueness has gained in importance as a result of the increasing interest in the approach of personalized medicine, new success rates in combating cancer and the numerous findings in oncological research. The scope of findings in this field will grow continuously. Therefore, it is important to standardize the basic documentation on data transmission and evaluation. Only a clear structure of the findings and the semantic uniqueness can contribute to better communication and to the improvement of the evaluation of this diverse information. But this aim requires that all above described conditions and procedures in genome analyzing are recorded and taken into account. In summary, the project's goals will only be realized with the cooperation of experts to provide evaluation and a quality assured mapping of the existing standards.

4 Conclusion

On the one hand, the standardization in the described field reduces the effort of documentation during the analysis in the laboratory. On the other hand, the semantically clear information can be searched more quickly and more precisely for appropriate therapeutic approaches and therapeutic

parameters. With this analysis, a finding can be made on the basis of which a personalized therapy can be created. This individual kind of treatment increases the chances of healing, relieves the side effects and generates medical knowledge. Through the optimization process, time resources in the laboratory as well as in the treatment can be performed more efficient.

By reducing the time needed, patients can benefit from a suitable therapy more quickly. Particularly in the case of cancer, the factor time is a decisive criterion for the success of a therapy. However, the optimization of the results cannot only save time. Publicly accessible databases can be browsed faster and more specifically with the given search masks, at least to contribute to better research requirements.

References

- [1] Federal Statistical Office Germany. The ten most frequent causes of death; 2013.
- [2] Barnes B. Report on cancer in Germany. Berlin: Robert Koch Institute; 2016.
- [3] Collisson EA, Cho RJ, Gray JW. What are we learning from the cancer genome? *Nat Rev Clin Oncol*.2012; 9:621–630.
- [4] Ullrich A, Sepulveda C, Yach D, et al. Public Health und Krebsprävention. *Der Onkologe* 2004; 10:166–174.
- [5] Krege S. Allgemeine onkologische Chemotherapie. In *Facharztwissen Urologie* 2014. Springer Berlin Heidelberg.
- [6] Seiffert S. Makromolekulare Chemie. *Nachr Chem*. 2016; 64:332–340.
- [7] Kuß A. Next Generation Sequencing. *medgen*. 2014; 26:229–230.
- [8] Bieber T, Broich K. Personalisierte Medizin. Zielsetzungen und Herausforderungen. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*. 2013; 56:1468–1472.
- [9] Kast K, Arnold N. Gendiagnostik bei familiären Tumoren. *Gynäkologe* 2012; 45:28–34.
- [10] Li M M, Datto M, Duncavage EJ, et al. Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer. *J Mol Diagn*. 2017; 19:4–23.
- [11] Helbig R. Prozessorientierte Unternehmensführung. Heidelberg: Physica-Verlag; 2003.
- [12] Töpfer A. Lean Six Sigma. Berlin Heidelberg: Springer; 2009.
- [13] <http://varnomen.hgvs.org/bg-material/simple/>
- [14] Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J, et. al. HGVS recommendations for the description of sequence variants: 2016 Update. *Human mutation*. 2016 ;37: 564-9.
- [15] Deckard J, McDonald CJ, Vreeman DJ. Supporting interoperability of genetic data with LOINC. *J Am Med Inform Assoc*. 2015; 22:621-7.
- [16] The International Organization for Standardization ISO/ TR 12300:2014 Health informatics - Principles of mapping between terminological systems. Geneva: ISO; 2014.