MEDINFO 2010
C. Safran et al. (Eds.)
IOS Press, 2010

doi:10.3233/978-1-60750-588-4-1334

The IT-Infrastructure of a Biobank for an Academic Medical Center

Andrea Dangl[a,b], Sara Y. Demiroglu[b], Jochen Gaedcke[a,e], Krister Helbing[b,c], Peter Jo[a,e], Fabian Rakebrandt[b,c], Otto Rienhoff[b], Ulrich Sax[a,b,c,d]

[a] *Clinical Research Unit 179, University Medical Center, Goettingen, Germany*
[b] *Department of Medical Informatics, University Medical Center, Goettingen, Germany*
[c] *Competence Network Congenital Heart Defects, Berlin, Germany*
[d] *Department of Information Technology, University Medical Center, Goettingen, Germany*
[e] *Department of General and Visceral Surgery, University Medical Center, Goettingen, Germany*

Abstract

For high quality research in biomedicine an operable biobank is essential. In order to make optimal use of the material and the huge amount of data a sustainable IT-infrastructure is indispensable. Therefore, we developed a concept for the IT-infrastructure of a biobank for an academic medical center. The conclusions for this concept are deduced from our experience with the biobank and IT-infrastructure of a clinical research unit. Our results indicate that the IT-infrastructure plays a pivotal role in successfully establishing a biobank. Several aspects of the IT-infrastructure are similarly found in other areas as, e.g. data protection and storage and quality management. Finally, we conclude that although a research database is not required for operating a biobank, the need for it will definitely emerge, especially with regard to personalized medicine and high-throughput gene expression analysis.

Keywords:

Biomaterial, Biobank, Biobank software, IT-infrastructure, Repository

Introduction

Biomaterial for research

Many medical research projects and clinical trials include endpoints requiring patho-histological reports, biomolecular tests and survival data. Therefore, it is indispensable to sample and store biological material to conduct clinical trials for finding biomarkers, improving existing, or developing new therapies, and to get statistically reliable results. Without enough high quality biomaterial many projects struggle or fail [1]. One solution to prevent the fast disappearance of biomaterial especially after the end of funding or in a follow-up phase, e.g. in a clinical trial, is a sustainable biobank (Fig. 1). Furthermore, providing a biobank with a solid IT-infrastructure increasingly becomes a criterion deciding about whether a grant application is accepted or not. It is worth to mention that in the beginning of a clinical trial the collection of adequate biomaterial is usually progressing slowly (not depicted in Fig. 1). A biobank should provide long-term storage of high quality biomaterial for research purposes. According to a definition of the German Ethics Council a biobank is the connection of a collection of human body materials and person-related data [2]. The biomaterial increases its value in combination with associated clinical patient data (MDAT) of the clinical trial. To administrate this huge amount of data a sustainable IT-infrastructure is indispensable.

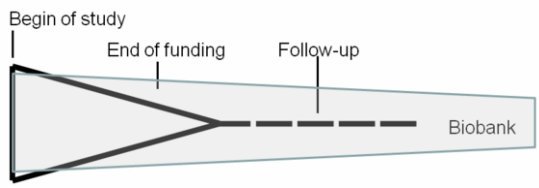

Figure 1- Availability of biomaterial during a clinical trial in comparison with a biobank

The biobank of the clinical research unit 179

In Goettingen the clinical research unit (KFO) 179 funded by the German Research Foundation (DFG) has the goal to develop an individualized therapy for the treatment of locally advanced rectal cancer. The standard treatment currently involves a preoperative radio-chemotherapy (RT/CT) followed by surgery, and an adjuvant chemotherapy. Beside the reduction of tumor size the striking advantage of preoperative RT/CT is the highly significant decrease of local recurrence [3]. Unfortunately, tumors respond very differently to the therapy ranging from complete response to resistance resulting in a survival benefit for only a subset of preoperatively treated patients [4]. Additionally, some patients suffer from severe side effects. For accomplishing their goal of a personalized medicine, the scientists perform experiments, like gene expression analysis, and promoter methylation analysis. For all these experiments they need high quality biomaterial of rectal cancer patients (normal and tumor tissue, as well as blood) that is taken at the initial staging examination of the tumor before

preoperative therapy is given. Moreover, blood samples are drawn during the therapy and of course, the resected cancer is formalin fixed and paraffin embedded to assess the response of the tumor to the neoadjuvant therapy and to finally determine the TNM staging. Thus, the biomaterial collected from rectal cancer patients within the KFO 179 is in the first line used to examine the biological basis of the individual tumor response in patients with rectal cancer. The material is then enclosed into the biobank and can be used by the researchers of the different sub-projects to answer their individual scientific questions. Having the administration data of the biomaterial and the MDAT in validated databases that are included in a sustainable IT-infrastructure has the advantage for the researchers to have an overview about the available material of each patient and the data that were already collected with this material.

Aspects to be considered when establishing a biobank

Although, biobanks are a good idea to counteract shortage of biomaterial, their establishment and maintenance should not be underestimated. There are a lot of aspects which need to be considered besides the storage of biomaterial. The most important aspects when establishing a biobank are in a chronological order:

1. the patient informed consent
2. ethical aspects
3. classification of the biomaterial by one single observer to avoid interobserver-variability
4. data security and protection
5. storage of biomaterial and its quality management
6. rules how to treat specimens with a different purpose of use
7. rules for the transfer of biomaterial to third parties

At the moment there are no biobank-specific laws in Germany, thus the work group biobanks of the telematic platform for research networks (TMF), consisting of scientists and researchers operating biobanks themselves, evolves guidelines, check lists, and privacy models for biobanks [5, 6].

Objectives

A medical center wide biobank would have the advantage that scientists with a specific research question requiring rare samples have a bigger chance to find sufficient samples to answer their question. Therefore, we want to develop a biobank concept with a focus on the IT-infrastructure for a whole academic medical center.

Materials and Methods

Approaching a biobank solution for an academic hospital

The concept for a biobank solution for an academic medical center is mainly based on our experience with the evolution of the biobank software from a remote data entry (RDE)-system for the KFO 179 biobank and supplemented by a literature research analyzing the requirements for larger biobank IT-infrastructures.

Developing a remote data entry system into a biobank software

Since 2007 several existing research projects affiliated with the Department of Medical Informatics at the University Medical Center in Goettingen recognized the need for a professional IT-support for their biomaterial collections. As a first effort, we developed a prototype system for the German Competence Network for Dementia based upon the existing remote data entry (RDE)-system (secuTrial). This product had several advantages over buying specific software, or completely developing software ourselves. Some of the advantages were: our RDE-system was validated in accordance with AMG/GCP, FDA 21 CFR Part11, good clinical practice (ICH-GCP), and TMF privacy models, the adaptation did not cause extra costs, and we were already experienced in setting-up clinical trials.

In 2007 the KFO 179 was founded. Due to its multitude of biomaterials to collect and several sub-projects involved, a web-based biobank software was needed. The continuation of the RDE approach allowed a role-based, audit trail equipped solution, despite the low IT budget of the project.

The KFO 179 biobank software in detail

Whereas the setup of the electronic case report forms for the clinical trials was a standard procedure, several challenges occurred setting up the biobank software. The existing functions of the RDE-system were improved according to the needs of the laboratory staff and a sample-oriented IT-based administration of biomaterial was developed, meaning that the data entry is barcode driven. The setup of the system required multiple redesign phases. Now, the system fully enables the monitoring of biomaterial quality in detail, documenting the storage, and capturing and documenting the respective usage. The basic data of the patients (IDAT) are stored in the study center. As the biomaterial collection does not only contain material and data from patients being treated in the University Medical Center, certain restrictions apply to the design of the IT-infrastructure. In order to fulfill the TMF privacy model, two separate software installations are required for the biobank [5]. One for the administration of the data of the biomaterial (BDAT) gained in the surgery and pathology departments and one installation for the pseudonymized MDAT, which are partially annotated by information from the hospital information system. Via an unambiguous barcode biomaterial specimens of patients can be linked with the information in the biomaterial database. The pseudonym of the patient can be linked to the MDAT in the second database only via an additional list, storing the pseudonym from the biomaterial database in connection with the pseudonym of the MDAT database. Both, BDAT and MDAT can be joined for research aspects via a mapping table. All results should finally be stored with a different pseudonym in a research database. De-identification of a patient is just possible in the study center keeping the IDAT.

The insights gained during this evolutionary process are very valuable for the design of a concept for larger biobanks.

Literature research for larger biobank IT-infrastructures

To identify the demands on biobank software for serving the purpose and fulfilling the TMF privacy model, a literature

research of requirements for biobank IT-infrastructures was performed by analyzing scientific literature and product information from providers of biobank information systems. Key words for the search in PubMed and google scholar were: biobank, biobanking, biorepository, IT-infrastructure, biobank software, biobanking software, BIMS, biorepository software.

Results

In our development process along with the literature research, we found the following points to be most crucial for an IT-infrastructure for a biobank.

Quality assurance

Quality assurance for biobanks has to take place on different levels: from monitoring of the samples to complete annotations during data entry into the database to Standard Operating Procedures (SOPs) describing what to do in case of a power failure.

Within the central laboratory of the KFO 179 the temperature of the samples is monitored manually. The temperature could also be manually entered into the RDE-system, but a direct communication between freezer and biobank software is not possible due to a lack of the necessary interface. That the temperature should be monitored was demonstrated in a study showing that a cyclic (5 times) temperature change from -196 to -150 °C reduces the viability of embryonic neurotissue cells to 50 % [7]. Having the appropriate interface, monitoring of single samples covering the temperature profile and freeze-thaw cycles over the whole life-span of a samples is possible[1].

In the KFO 179 data are entered into the RDE-system by the laboratory personnel for each single sample comprising two entry steps. In a first step, the barcode, the name of the person taking the sample, the date when the sample was taken, the patient`s pseudonym, and the sample`s origin are entered. After the isolation of e.g. RNA from this sample and its quality assessment (taking place just in the central laboratory), the RIN (RNA integrity number) describing the degradation state of the RNA [8], and other quality parameters are entered into the respective form in the RDE-system. Moreover, the storage location, the description of the type of isolated material, and the exact amount and concentration are documented. If a scientist wants to use the same sample for research, the RIN gives information, whether the sample could be used for a certain research purpose. The workflow of the KFO 179 is in accordance with literature mentioning that before samples are included into the biobank, a quality assessment should take place to determine the purity and degradation of e.g. tissue samples [9].

The RDE-system used by the KFO 179 is equipped with an audit trail and a role-based access control. On the data level a check for plausibility such as defined ranges for distinct values exists. Due to the structure of the software (originally designed for clinical trials) based on visit plans, it is not possible to determine whether a single storage location for biomaterial is entered twice, neither to allocate free storage places nor to give an overview about the storage places and their capacities. An identification of duplicates can just be performed after the export of the data and an analysis with statistical analysis system (SAS). In comparison to the literature, the RDE approach has some disadvantages[1].

In addition to high quality biomaterial, high quality annotations are also required. It must be documented what type of biomaterial is stored (using, e.g. the ICD classification), when and with which method it was extracted, purified, and preserved, and how the material was stored during its life-cycle.

In the central laboratory of the KFO 179 an alarm system is installed. In case of a power failure the laboratory personnel follows an emergency plan including a defined workflow for saving samples. This KFO 179 concept is in accordance with literature describing the need to set up a concept in the form of SOPs describing what to do in events of power failure or broken freezer systems [10].

Document and workflow management

The KFO 179 uses a web-based portal system run by the Department of Medical Informatics to save essential documents like study protocols or SOPs. These documents are stored in distinct folders for each sub-project. In addition, the portal offers access to several applications so that the progress can be monitored and collaborations can be fostered. It is described that certain laboratory information management systems (LIMS) are able to directly provide such document storage services without the need for an additional portal[1].

Search functions

In order to allow long-term sample scheduling, researchers need a permanent overview of samples present in the biobank.

Within secuTrial it is possible to get overviews of the biomaterial present in the biobank including a sorting function. For more complex queries, the data needs to be exported and analyzed with SAS. The same applies to the statistical analysis after the end of the study. Some LIMS[1,2] allow a detailed query to work with the biomaterial in the biobank.

Sample request and distribution

One function of biobanks is the storage of biomaterial. Therefore, an essential aspect is how to deal with sample requests and how to serve them.

Within the KFO 179 a transfer of biomaterial to the single subprojects works as follows: A scientist asks for biomaterial in the central laboratory and gets it and the transfer is manually recorded in the RDE-system. The access to MDAT is controlled by a data management and safety committee (DMSC). A transfer of biomaterial to third parties is not planned so far. Our literature analysis revealed that a biobank should offer a web-based possibility to request, and if possible, also to search for samples [10]. In this way, the whole process could be documented at the same time. To minimize errors in issuing samples to third parties and to be in agreement with data protection laws, an IT-supported identification method like RFID

[1] See: http://www.starlims.com/Biobanking_Brochure_final.pdf
[2] See: http://www.biofortis.com/products/labmatrix

tubes or the more common barcode tags should be used on each sample instead of identifying data of the patient. The transfer of MDAT and biomaterial to third parties should just take place in anonymized form [5].

Development of a concept for an IT-infrastructure for an academic medical center wide biobank

Covering the aspects to be considered when establishing a biobank (see introduction), we can deduce the following concept for an academic medical center considering the KFO 179 solution and the literature research:

(1.) The usage and transfer of samples have to be described very detailed in the patient informed consent. The patient has to be informed, whether the sample and the generated data are used only in a specific scientific area, e.g. cancer research, or generally for research purposes. Moreover, it must be possible for the patient to withdraw the consent at any time.

(2.) Ethical aspects: A DMSC must be installed to preserve the safety of patients. In addition, it secures the reliability of the data, which has to be guaranteed, and the transparency of the working processes within the projects.

(3.) To avoid interobserver-variability, all observers have to synchronize their classification methods. Within a clinical trial there should be only one single observer examining the biomaterial and for example staging it. A central image-bank could be one solution to solve this problem.

(4.) Data security and protection: Role-based access control, encrypted transfer of data into the RDE-system, separate storage of MDAT, IDAT, and BDAT, and pseudonymization of the data must be given. The IT-infrastructure must be setup according to data protection laws, TMF guidelines [5], and requirements of the ethic committee of the respective academic medical center.

(5.) Quality assurance must take place on all levels. Especially for a larger biobank it is necessary to ensure high quality biomaterial as well as very detailed annotations. The temperature profile of each sample should be monitored. Thus, an increase in temperature due to a defect or a power failure could directly lead to a warning message to the responsible persons and counteractions could be taken immediately. Unless one directly uses room-temperature sample storage for DNA or RNA obviating the need for freezers [11]. The entry of data into the biobank software must be controlled by a role-based access control and the values added into the fields of the forms should have plausibility checks also for duplicates of barcodes or storage locations. The data entry into a biobank software depends on the state of automation of the biobank, i.e. whether robots are used for storage locations or making aliquots of samples [12].

(6.) It is necessary to store biomaterial of clinical trials for a long time. The progress of research is very fast and therefore it is important that researchers can use datasets of former trials e.g. to establish and validate new methods. In the KFO 179 such an older comparable dataset is used for examination aspects in order to reproduce the results. In case of different results it might be helpful to generate data of older samples again using the current technology. For reducing the problem of treating specimens stored with a different purpose and to label them and determine their content, it is necessary to anonymize them. Nevertheless, for a complete new biobank we recommend to first setup a functional IT-infrastructure including machines for reading samples, labeling, and storing them. Just upon the complete existence of the framework a biobank should start collecting samples.

(7.) A web-based tool for third parties to enquire biomaterial is essential. For the transfer of biomaterial to third parties a DMSC is required, controlling the giving away and required measures like anonymization. An interesting solution is practiced at the biobank of the Medical University in Graz, Austria. There, third parties are supported by the biobank to write research proposals. The biobank determines the availability of the required samples and performs all experiments within core facilities of the Medical University. Only the research result is given away to the third party upon payment[3]. Thus, valuable samples are not given away and a quality loss of the specimen caused by an interference of the cold chain can be excluded.

Finally, it is indispensable to avoid parallel solutions meaning that per academic medical center just one biobank should exist. Most importantly we recommend a new biobank software for an academic medical center, as the secuTrial approach has too many limitations and is not sufficient anymore.

Discussion

Our experience showed that around younger scientists there is a big acceptance of new ideas, whereas senior laboratory personnel tend to generally dislike changes. Nevertheless, existing skepticism is quickly released when the personnel realizes the simplifications, which such new ideas bring with them for the daily workflow.

On the one hand, compared to a distributed storage approach one central biobank has the advantage of saving costs regarding cooling facilities, alarm, and power back-up systems [10]. Optimally, the biobank should be linked to the central laboratory for health care to take over otherwise discarded samples. On the other hand, room must be created for many samples and it requires an elaborate workflow to secure continuous cooling of the specimens from the place of retrieval to the biobank freezers. Moreover, property issues and data sharing issues can easily inhibit the central approach. Therefore, a mature biobank software should not be depending on local issues.

The financial aspect of the biobank depends on the state of automation, e.g. machines for aliquoting, refractioning of blood, storage of samples, labeling of samples, but also on the software for the IT-infrastructure and the number of employees for the biobank.

An academic medical center wide biobank offers advantages and disadvantages to its scientists, although the advantages prevail. The main advantage is that scientists can revert to a larger selection of samples for their research. The fact that

[3] See: www.meduni-graz.at/1449

physicians would need to hand in all their samples gained from patients not used for medical examination into the biobank might be a disadvantage on the first sight. In reality this is an advantage, as it allows the blinding of the physician, who might have a double-role as clinician treating patients and scientist working with their material.

In the KFO 179 an emerging need for an easy-to-use research database (RDB) came up recently. So far, this lies beyond the possibilities of an RDE system. A suggestion for an IT-infrastructure for an academic medical center including a RDB is described in figure 2 [13].

The MDAT stored in the study database (SDB) are encrypted with a person identifier (PID), as well as the IDAT which are stored in a separate list. The BDAT in the biomaterial database (BDB) are organized via a laboratory identifier (LabID) and have a different patient pseudonym. Before researchers can start working with their data, e.g. combining gene expression results with MDAT, a mapping of the identifiers must take place preferably by IT. A RDB into which the matched datasets from SDB and BDB and genetic data could be imported via several interfaces could be searched by a query tool like i2b2 (informatics for integrating biology and the bedside)[4] [14, 15].

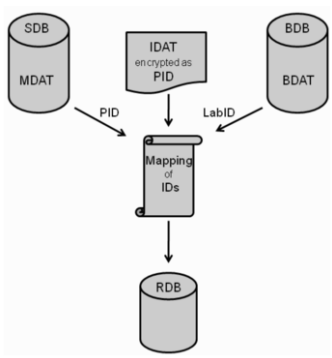

Figure 2- IT-infrastructure for an academic medical center in accordance with German data protection laws

Conclusion

A biobank for an academic medical center should be developed for every university paying special attention to the IT-infrastructure. The choice of software for the single components (Fig. 2) should be well-thought of as the IT plays a pivotal role. The need for a RDB will just slowly arise during operating the biobank, but the person in charge should at an early stage decide how a product could fit best into the existing IT landscape.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (KFO 179).

[4] See: https://www.i2b2.org

References

[1] Waltz E. Pricey cancer genome project struggles with sample shortage. Nat Med. 2007: 13 (4): 391.

[2] Ethikrat. Biobanken für die Forschung - Stellungnahme. Berlin: Saladruck; 2004.

[3] Sauer R, Becker H, Hohenberger W, Rodel C, Wittekind C, Fietkau R, Martus P, Tschmelitsch J, Hager E, Hess CF, Karstens JH, Liersch T, Schmidberger H, Raab R. Preoperative versus postoperative chemoradiotherapy for rectal cancer. N Engl J Med. 2004: 351 (17): 1731-40.

[4] Rodel C, Martus P, Papadoupolos T, Fuzesi L, Klimpfinger M, Fietkau R, Liersch T, Hohenberger W, Raab R, Sauer R, Wittekind C. Prognostic significance of tumor regression after preoperative chemoradiotherapy for rectal cancer. J Clin Oncol. 2005: 23 (34): 8688-96.

[5] Simon JW, Paslack R, Robienski J, Goebel JW, Krawczak M. Biomaterialbanken - Rechtliche Rahmenbedingungen. Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft, 2006.

[6] Kiehntopf M, Böer K. Biomaterialbanken - Checkliste zur Qualitätssicherung. Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft, 2008.

[7] Vysekantsev IP, Gurina TM, Martsenyuk VF, Petrenko TF, Kudokotseva EV, Koshchiy SV, Groshevoy MI. Probability of lethal damages of cryopreserved biological objects during storage. Cryo Letters. 2005: 26 (6): 401-8.

[8] Mueller O, Lightfoot S, Schroeder A. RNA Integrity Number (RIN) - Standardization of RNA Quality Control. Agilent Technology; 2004 [Access date: 14.10.2009]; Available from: http://www.genome.duke.edu/cores/microarray/documents/Using%20RIN.pdf.

[9] Shevde LA, Riker AI. Current concepts in biobanking: development and implementation of a tissue repository. Front Biosci (Schol Ed). 2009: 1: 188-93.

[10] Riegman PH, Dinjens WN, Oosterhuis JW. Biobanking for interdisciplinary clinical research. Pathobiology. 2007: 74 (4): 239-44.

[11] Blow N. Biobanking: freezer burn. Nature Methods. 2009: 6 (2): 173-7.

[12] Blow N. Lab automation: tales along the road to automation. Nature Methods. 2008: 5 (1): 109-12.

[13] Simon J, Paslack R, Robienski J, Cooper DN, Goebel JW, Krawczak M. A legal framework for biobanking: the German experience. Eur J Hum Genet. 2007: 15 (5): 528-32.

[14] Murphy SN, Mendis ME, Berkowitz DA, Kohane I, Chueh HC. Integration of clinical and genetic data in the i2b2 architecture. AMIA Annu Symp Proc. 2006: 1040.

[15] Mendis M, Phillips L, Kuttan R, Pan W, Gainer V, Kohane I, Murphy SN. Integrating Outside Modules into the i2b2 Architecture. AMIA Annu Symp Proc. 2008: 1054.

Address for correspondence

Dr. Sara Y. Demiroglu, University Medical Center Goettingen, Department of Medical Informatics, D-37099 Goettingen, Germany.
Email: sara.demiroglu@med.uni-goettingen.de