Supplementary 'biobank ID implementation example' document

The following example from a major biomedical research project based in Germany illustrates the application of the how-to-guide to design biospecimen IDs and how to fully integrate this and the accompanying functions into the biospecimen management software of a biobank.

Description of the Biobank

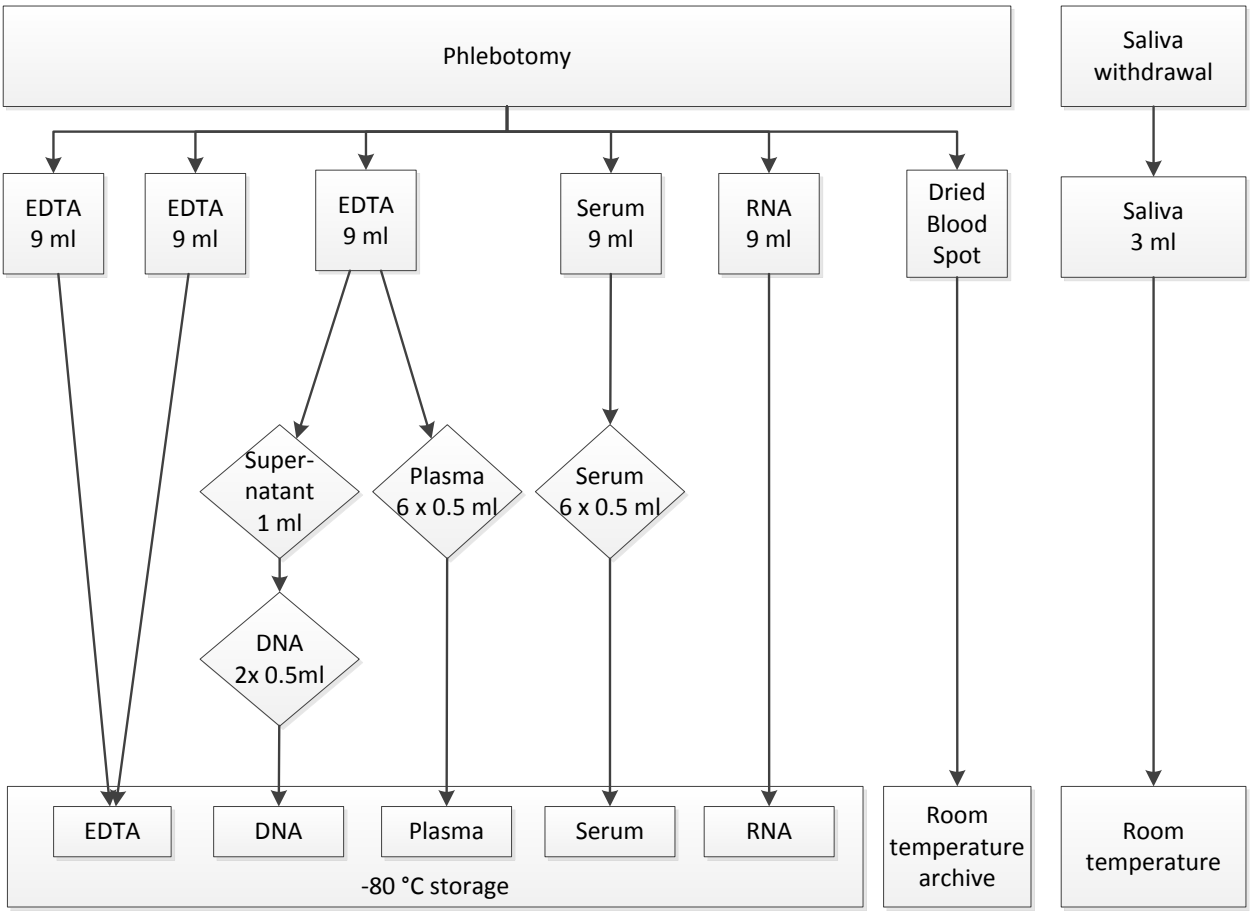
The biobank belongs to a clinical research group named KFO 241 "Genotype-phenotype relationships and the neurobiology of the longitudinal course of psychosis". This biobank according to Watson and Barnes (1) is a mono-user biobank only serving this research project. It is large in size (> 1000 cases) as samples from 6200 cases (patients and healthy controls) will be collected at four consecutive time-points within 18 months. As many as 20 collaborating centers from northern Germany will participate in recruiting donors. All biospecimens and data will be stored centrally in Göttingen.

The biospecimen management software selected is part of a larger IT-infrastructure (2) and is a commercially available product. Changes to this software can either be made by the vendor company or by the IT department responsible for running this application.

Analysis of the Biobank's Workflows

In general, each study participant should have four encounters with the personnel of the research project. At each encounter clinical data as well as five primary tubes of blood or one tube of saliva will be collected from each participant. A detailed analysis of the workflows yielded three sub-processes. These are the recruitment of the study participant, the phlebotomy and processing of blood into derivatives, and the interview with the study participant. The sharing of samples within the research project will only take place at a later stage. The only process related to biospecimen handling here is the second one. Thus, for this paper we focused in detail on this particular workflow, which is depicted in Supplementary Figure 1. The process of withdrawal and processing of biospecimens starts with phlebotomy to obtain a

dedicated biobank biospecimen during routine health care visits or in the study office (healthy volunteers). The samples are then transferred to the research lab where they are being processed further (see Figure 1). Finally, biospecimens are stored as frozen samples, and relevant data is entered into the professional biobanking management software. The details of processing the biospecimen in the research lab are depicted in Supplementary Figure 1.



SUPPLEMENTARY FIGURE 1 PROCESSING SCHEME FOR THE BIOSPECIMEN SAMPLES OF THE KFO 241

At each visit, the study nurses collect five different tubes of blood from each study participant with different fixatives: 3x EDTA blood (each 9 ml), 1x serum (9 ml), and 1x blood for RNA isolation (9 ml) for the isolation of RNA. During the processing of blood samples in the lab, up to 12 0.5 ml aliquots of

plasma and serum are going to be generated from two blood tubes collected from each study participant. In case the study participant does not want to or cannot donate blood, saliva will be collected in a dedicated tube for DNA isolation out of saliva instead. In addition, a dried blood spot paper is stored as a back-up DNA reference in the rare event that the label of the blood samples is lost.

Biospecimens will be stored in the biobank of the Department of Psychiatric Genetics of the University Medical Center in Göttingen.

Requirements Analysis

Requirements were gathered from the stakeholders, existing documents and existing systems. The relevant stakeholders involved in this biobank example were the lab and research personnel of the KFO 241 responsible for working with the biospecimen as well as the IT personnel administrating the IT-infrastructure. Documents found to be important were the ISBER Best Practice Guidelines (3) as well as the German Data Protection Concept for networked medical research projects (4). Further requirements were imposed by the biospecimen management software as well as the fact that this research project is a multi-center project with a centralized biobank and a centralized web-based biospecimen management software.

Altogether, the requirements could be distinguished into three categories: requirements related to the organization of the research project (O; see Table 1), requirements related to the identification of biospecimen (I; see Table 2), and requirements related to the management of biospecimen information in the software (M; see Table 3).

Three requirements related to the general organization of a research project were identified. Seven requirements regarding the identification of biospecimens were collected. The requirements can be subdivided into requirements describing the ID for biospecimen (I.1-I.6), and the visualization of the ID (I.3 and I.7).

For the administration of biospecimens in a biospecimen management software eight requirements were identified. The requirements can be grouped into sample-specific (M.1-M.3) and general requirements (M.4-M.7). It must be ensured, that the implementation of the above stated requirements does not influence existing interfaces of the biospecimen management software with other components of the IT-infrastructure (M.8) of this research project as described elsewhere (2).

Developing the Concept for Biospecimen IDs and Full Integration into a Biospecimen Management Software: the Kit Concept

a) Concept for Biospecimen IDs

We named the concept, which we developed following the how-to-guide as described in the source document, the kit concept. In this concept a distinct set of tubes – a kit - is used per study participant and per visit. All biospecimen containers should be from the same company including the same additives. Each kit has a unique kit-ID. The kit-ID is a unique string with a consistent structure as follows: The prefix *Kit*, *a project ID automatically set per project*, and a *unique random 10-digit number* generated by the biospecimen management system. The unique number is generated with a java random number generator. The result of the random number is multiplied by 10^{10} to get a 10 digit number, where all decimal places are eliminated. For each new number it is verified that it does not already exist in the system. If so, a new number will be generated. If not, the random number will be used as part of the ID.

Kit_projectID_1234567890

Each tube in a kit is assigned a sample-ID consisting of the kit-ID plus a 3-letter code for the type of biospecimen (MatCode; e.g., BLD for blood) and an ascending number starting from 1. The 3-letter-code for the type of biospecimen is based on the 3-letter-code of the biospecimen type as the first element of

the Sample Preanalytical Code (SPREC; (5,6)). With the ascending number at the end each sample will be uniquely identifiable.

Kit_projectID_1234567890_MatCode_AscendingNumber

Depending on the needs of a biobank or research projects, different kits can be created in the biospecimen management system and the corresponding labels can be printed and attached to the respective tubes. Examples of different kits taken from our experience include a standard kit for blood withdrawal in routine health care or dedicated rooms, an extension kit for sample processing in the lab, and an external kit for collaborating partners without a fully equipped facility to process the samples on-site.

The sample ID can be printed onto adhesive labels, where they are visible as plain text in the format described and depicted above and as a machine-readable 2D barcode. We chose the data matrix 2D barcode format for that purpose as it is possible to print this kind of 2D barcode in a very small format. Therefore very small aliquots can be chosen for containers.

In practice, every kit is packed into a re-sealable specimen transport bag labeled with the kit-ID (human and machine (2D barcode) readable. The standard kit of the KFO 241 contains a butterfly connected to a tube, an adaptor to the tubes, five tubes for the collection of blood (acting by using vacuum for filling), and one archive paper for a dried blood spot (see Supplementary Figure 2).

SUPPLEMENTARY FIGURE 2 KIT (SPECIMEN TRANSPORT BAG WITH EMPTY SAMPLE CONTAINERS AND PHLEBOTOMY TOOLS AS WELL AS ADDITIONAL PAPER SHEETS IN THE SLEEVE)

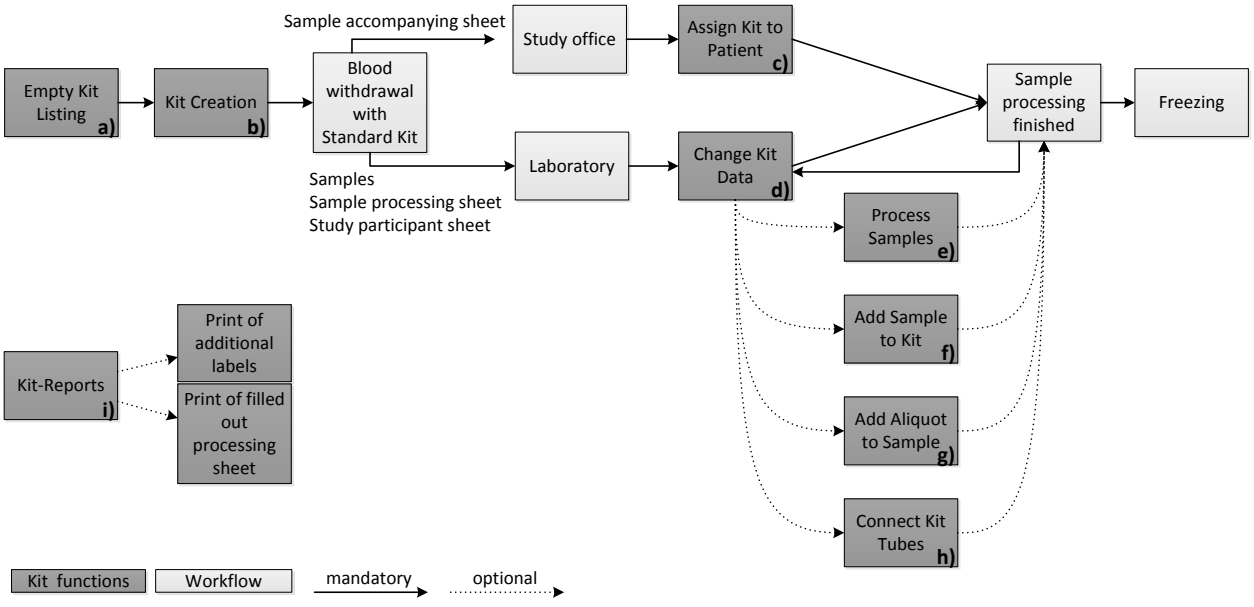
Applying the kit concept labeling scheme to the above described standard kit with its containers is depicted in the main manuscript in Figure 2. Note that as SPREC Version 1 (5) did not have a 3-letter-code for dried blood spot, we generated our own code for that (AAA).

The back of each re-sealable specimen transport bag contains a small sleeve, where accompanying sheets of paper could be stored. In our concept, each kit additionally contains three sheets of paper: the study participant sheet, the sample accompanying sheet and the sample processing sheet. The kit-ID is affixed to the study participant sheet, where the identifying data (e.g., name, birthdate, and address) of the participant is collected. The sample accompanying sheet also includes the attached kit-ID and information regarding the phlebotomy (e.g., phlebotomist, withdrawal date and time). The sample processing sheet is generated by the biospecimen management system and already contains the kit-ID as

well as all sample-IDs with the corresponding 2D barcodes. Here, handwritten notes can be added for each sample, i.e. SOP/protocol deviations, which need to be entered into the biospecimen management system later on. It would also be possible to include SOPs related to the kit within the small sleeve. This could be very useful for external study centers that will not collect biospecimen on a regularly basis. With the kit-concept explained so far, all of the requirements from Table 2 regarding the biospecimen ID are fulfilled.

b) Full integration of Biospecimen-ID related features into the biospecimen management system: How the developed concept will support the biospecimen workflow from an IT-perspective

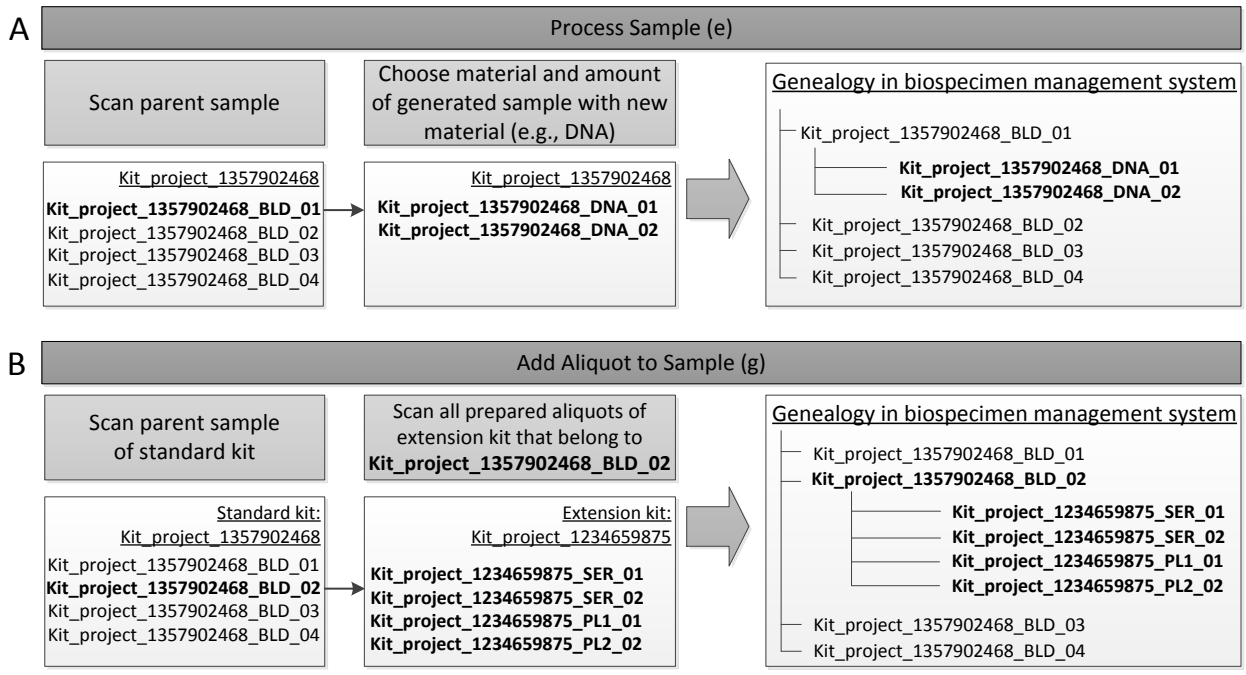
In order to explain how the standard biobanking workflow is supported using the kit concept, it is necessary to depict the workflow of collecting and processing biospecimen along with the respective implemented function of the kit concept (Supplementary Figure 3).



SUPPLEMENTARY FIGURE 3 SUPPORT OF BIOBANKING STANDARD WORKFLOW USING THE KIT CONCEPT FUNCTIONS. THE WORKFLOW FROM WITHDRAWAL OF BIOSPECIMEN TO SAMPLE PROCESSING AND FREEZING IS DEPICTED (WHITE BOXES). THE FUNCTIONS IMPLEMENTED FOR THE KIT CONCEPT ARE DEPICTED IN GREY BOXES. MANDATORY STEPS OF THE WORKFLOW HAVE ARROWS, OPTIONAL STEPS HAVE DASHED ARROWS.

Supplementary Figure 3 shows the whole workflow from creating a kit to freezing the biospecimens. As a first step of the workflow, users can check how many available kits exist of each kit type, using the "Empty Kit Listing" function (a). If insufficient empty kits are available, new kits can be generated using the "Kit creation" (b) function. This function generates empty kits in the biospecimen management system including a kit-ID and respective samples-IDs for the pre-selected numbers and types of biospecimens. It allows printing of kit-IDs and sample-IDs and the sample processing sheet. The adhesive labels with the sample IDs should be applied onto the respective tubes, the kit-IDs should be applied to the sample accompanying sheet, the study participant sheet, and the specimen transport bag and all items should be packed together. Afterwards, the pre-labeled kits can be sent to external collaborating centers for use. These steps fulfill the organizational requirements that all centers use the same labeling system (O.1) and that the kits can be packed in advance (O.2). Following the step of kit creation in the biospecimen management system, the biospecimen can be collected and the respective documentation added to the sample accompanying sheet (sample data) and into the study participant sheet (study participant data). Afterwards, the kits are sent back for central processing and storage, and the study participant sheet is forwarded to the study office, and the samples, the sample processing sheet, and the sample accompanying sheet are forwarded to the lab. From this point in time on, the biobank support workflow is split (see Supplementary Figure 3). In the study office, the kit can be linked to a participant's ID using the "Assign Patient to Kit" (c) function. In order for this to work, the staff of the study office requires the identifying information of the study participant from the study participant sheet to generate a participant ID as described elsewhere (2). The participant ID can be printed afterwards and is stored in the research participant's study file. By scanning the participants ID and the corresponding kit-ID into the respective fields within the "Assign Patient to Kit" function, all samples of this kit will be automatically linked to the participants ID. This function covers the organizational requirement that any kit can be used for any study participant (O.3). In parallel, in the laboratory the biospecimens are either directly stored or being processed further and then stored. The storage location can be either directly documented in

the IT tool by simply scanning the unique box-ID and choosing the respective storage location within this box for each sample or by documenting the storage location on the sample processing sheet. The latter would require a later manual entry of the storage location of the sample into the biospecimen management system. Any deviations from standard operating procedures (SOPs) should be documented onto the sample processing sheet for the respective sample. Using the function “Change Kit Data” (d), either relevant data for all samples of a kit can be changed simultaneously (i.e. date of phlebotomy) or information for each individual sample can be edited (i.e. volume or pre-analytical variables (5,6)). For the KFO 241, the processing of samples in the laboratory requires the use of an extension kit. The extension kit, which are only required in the lab for processing of samples, holds 12 smaller sample containers for the storage of plasma and serum aliquots. During the creation of extension kits, the kit-IDs and sample-IDs are printed but will not be attached to the containers straightaway, as the filling height of the 10 ml blood sample containers might differ and thus will not always yield the same number of aliquots. Thus, no containers will be wasted by pre-sticking sample-IDs on them. If the number of containers and corresponding sample-IDs within an extension kit are not sufficient, the function “Add Sample to Kit” (f) can be used. This function allows the addition of further sample containers with sample-IDs fitting to an existing kit. Once, the aliquoting step is finished, aliquots of an extension kit (different MatCode according to (5,6)) can be assigned to their respective parental sample using the “Add Aliquot to Sample” (g) function. As depicted in Supplementary Figure 4B the kit-ID of the extension kit differs from the kit-ID of the standard kit. However, the assigning of children samples to parental samples from then on will always be visible in the genealogy tab of the selected sample. For sporadically occurring events, e.g. DNA isolation, and where no special kit was designed for that purpose the “Process Sample” (e) function can be chosen allowing the specification of the number of aliquots and the choice of the type of biospecimen. The newly generated sample-ID for the DNA sample would have the sample kit-ID as its parental sample but a different MatCode (see Supplementary Figure 4A).



SUPPLEMENARY FIGURE 4 CHANGING OF THE MATERIAL TYPE OF A BIOSPECIMEN DEPENDING ON THE FUNCTION USED. USING THE PROCESS SAMPLE FUNCTION LEADS TO ALIQUOTS WITH THE SAME KIT-ID AS THE PARENTAL SAMPLE, WHEREAS USING THE ADD ALIQUOT TO SAMPLE FUNCTION ALLOWS TO CONNECT SAMPLES WITH DIFFERENT KIT-IDS.

If for the aliquotation process, tubes with a laser-etched 2D barcode were used, the function “Connect Kit Tubes” (h) enables the user to permanently link the laser-etched 2D barcode at the bottom of a container with the system generated sample-ID attached to the same tube. At any time the “Kit-Reports” (i) function can be used to reprint any report or ID.

Not all functions are dependent on each other: Assign kit to patient (c) and Change kit data (d) do not have to take place in parallel or in a special order. The functions Empty kit listing (a), Kit-reports (i), Process samples (e), Add sample to kit (f), Add aliquot to sample (g), and Connect kit tubes (h) can be used at any time and e-h) are not mandatory.

The above described functionalities (a-i) of the kit concept fulfill all requirements regarding the management of biospecimen within the biospecimen management system as depicted in Table 3.

Implementation and Validation of the Concept

A summary of the different functions implemented by the KFO 241 into their biomaterial management system is given in Supplementary Table 1.

SUPPLEMENTARY TABLE 1 DETAILS ON THE TEN KIT CONCEPT FUNCTIONS ARE GIVEN IN CHRONOLOGICAL ORDER ACCORDING TO THEIR APPEARANCE IN THE BIOBANKING WORKFLOW AS DEPICTED IN SUPPLEMENTARY FIGURE 3.

Name of the kit concept function	**Description of the function**
Administrative Function:	
Kit Manager	The administrator can define different kit types per project and assign each kit a distinct number of tubes with specified material types. For each sample its dedicated sample container has to be chosen (e.g. EDTA, serum and a fully customizable metadata template can be assigned or newly generated. It is possible to assign kits to specific projects to prevent users from having access to all existing kit types throughout the biospecimen management system. Furthermore, it is possible to use standard kit types throughout the biospecimen management system and adapt them for special needs in a given project without creating a totally new kit type.
User functions:	
a) Empty Kit Listing	Users can monitor the number of empty kits of each type available within their project. If an insufficient number of empty kits are left, new ones should be created.
b) Kit Creation	The user can choose the type of kit to be created in the biospecimen management system (from a list generated by the administrator in the "Kit Manager"). With choosing a kit type, IDs and the corresponding sample processing sheet with all IDs on it can be printed. At the same time all samples of a kit are created as empty entries in the software. At this point of time, the empty kits are not assigned to any participant. Thus, it is possible to assemble kits in advance and send them to external collaboration partners. As mentioned in a) users only have access to kit types that are listed for their project and are not able to create kit types on their own to prevent mixed up kit types.
c) Assign Patient to Kit	Kits are created with empty sample containers in the biospecimen management system and without linkage to any study participant, as described above. Thus, at one point dedicated staff has to connect a kit with the participants' ID. This can be done at any time in the entire workflow. The linking should be done by a person, e.g. the study nurse of the study office, allowed to see the identifying data of a study participant due to German data privacy regulations. The kit-ID needs to be scanned into the respective field of a form and the study nurse either has to request a participants ID with the information from the study participant sheet (2) or only needs to

	scan the existing participant ID into the respective field in the biospecimen management system. With this step all samples of a kit are assigned to a unique participants' ID in one step.
d) Change Kit Data	The users can change and edit data and metadata of kits and their included samples. Metadata could be e.g. the person responsible for the withdrawal of a biospecimen or the withdrawal time. After scanning a kit-ID the user can change the following information: Metadata of the kit itself, storage location of the samples, volume of samples, metadata of samples, and the pre-analytical variables, i.e. SPREC (5,6) of samples.
e) Process Sample	The "Process Sample" function allows the documentation of sample processing, e.g. for DNA extraction the original samples have to be scanned and the desired number of new aliquots has to be chosen. The biospecimen management system will generate the respective sample-IDs for the new DNA samples with the corresponding material type fitting the kit-ID. The processing is not limited to DNA extraction but fully customizable to fulfill all needs.
f) Add Sample to Kit	If for any reason more sample containers together with sample-IDs are needed (e.g. due to natural variation or a better yield), the "Add Sample to Kit" function allows the user to add further containers of any type to an existing kit. After choosing the required container and the type of biospecimen the corresponding sample-IDs and an updated sample processing sheet can be printed. The sample-ID for the additional container consists of the same kit-ID as the existing samples with incremented counter at the end.
g) Add Aliquot to Sample	When a biospecimen is processed, e.g. a blood sample of the standard kit, and the resulting plasma is aliquoted into tubes of the extension kit, the "Add Aliquot to Sample" function allows the linkage between the sample-IDs of the parental blood sample and the children plasma aliquots. Users have to scan the barcode of the respective parental sample of the chosen standard kit, which was processed. Afterwards, all aliquots (sample-IDs of the extension kit) have to be scanned and the volume of each sample has to be documented. The connection is now visible in the tree-like genealogy of the samples.
h) Connect Kit Tubes	Besides the 2D data matrix barcode generated by the biospecimen management system some tubes have 2D barcodes laser-etched at their bottom. This laser-etched barcode can be registered in the biospecimen management system as a fallback in case the adhesive label falls off. For this, the "Connect Kit Tubes" function has to be used. Users have to scan the kit-ID to get a listing of all samples in this kit. Subsequently, they have to scan the system-generated sample-IDs followed by the laser-etched tube barcodes. From then on both 2D barcodes are connected, known, and searchable in the system.

i) Kit Reports	The user can choose between different types of reports to be printed at any time. This could either be ID-labels including a 2D data matrix barcode (kit-ID, sample-ID, or participants' ID) or information sheets about the samples. The information sheets include the sample processing sheet or a report filled with information from the database on all samples of a kit.

We were in close contact with the vendor throughout this process - from ordering the required function via a contract up to the handover of the developed data packages – so we were able to intervene immediately if any functions (in particular user friendliness) did not follow our requirements.

After receiving the different software packages from the programmer of the biospecimen management software we validated the software packages. All bugs were reported back and revised software packages were retrieved.

Finally, before the users could play in the test system of the biospecimen management system a training session was provided by dedicated IT-personnel and an instruction manual designed by us was handed over. This instruction manual described in detail the different functionalities implemented in the software and contains a lot of screenshots of the system for better understanding and conformability.

Footnotes

Requirements Analysis

Although suggested in the how-to-guide, we did not prioritize the requirements, as for example lined out in the IEEE standard (7) as funds for adaption of the biospecimen management software were available and thus all of the requirements could be implemented.

Satisfaction of the users

At the end of November 2013 11100 samples (roughly 3000 blood, 3000 plasma, 3000 serum, 1700 DNA, and 400 dried blood spots) in 1115 kits from 410 study participants were documented in the biospecimen management system of the KFO 241. Moreover, our example solution allows high quality annotations of biospecimens while minimizing the documentation workload for the researchers.

Overall, the users of the KFO 241 are very satisfied with the biospecimen management system and the way their biospecimen IDs are generated as observed and asked with informal questions. There were no concerns related to data entry functions or biospecimen labeling, nor did researchers, information technologists or the ethics committee express concerns. Minor issues related to the kit concept were always related to human errors. They were immediately reported and could be fixed, e.g. samples were disposed too early in the workflow or unused kits needed to be deleted as some of the containers were beyond their expiry data already.

In consulting potential future users of this concept (e.g. those planning the Brain Bank of the Competence Network Multiple Sclerosis, www.kompetenznetz-multiplesklerose.de, and those planning the university medical center-wide biobank in Göttingen) an adaptation that would allow the user to generate kits on the fly by only scanning the 2D barcodes of the sample containers at the bottom and selecting the respective biospecimen type in the system has been suggested. This is only one possibility for further kit concept development.

The biospecimen ID generation as described above is generic and therefore can be applied to many projects and is probably done in a similar fashion already. As most biomedical research projects collect a defined set of specimen at a given time point, we think, the usage of the kit concept will find broad acceptance.

References

1. Watson PH, Barnes RO. A Proposed Schema for Classifying Human Research Biobanks. Biopreservation and Biobanking. 2011;9(4):327–33.

2. Demiroglu SY, Skrowny D, Quade M, et al. Managing sensitive phenotypic data and biomaterial in large-scale collaborative psychiatric genetic research projects: practical considerations. Mol Psychiatry. 2012;17(12):1180–5.

3. International Society for Biological and Environmental Repositories. 2012 Best Practices for Repositories - Collection, Storage, Retrieval, and Distribution of Biological Materials for Research. Biopreservation and Biobanking. 2012;10(2):79–161.

4. Pommerening K, Sax U, Müller T, et al. Integrating eHealth and Medical Research: The TMF Data Protection Scheme. eHealth: Combining Health Telematics, Telemedicine, Biomedical Engineering and Bioinformatics to the Edge. 1st Ed.. Berlin: Aka; 2008. p. 5–10.

5. Betsou F, Lehmann S, Ashton G, et al. Standard preanalytical coding for biospecimens: defining the sample PREanalytical code. Cancer Epidemiol Biomarkers Prev. 2010;19(4):1004–11.

6. Lehmann S, Guadagni F, Moore H, et al. Standard Preanalytical Coding for Biospecimens: Review and Implementation of the Sample PREanalytical Code (SPREC). Biopreservation and Biobanking. 2012;10(4):366–74.

7. IEEE Computer Society. IEEE Std 830-1998: IEEE Recommended Practice for Software Requirements Specifications. IEEE; 1998.