Master of Science in Informatics at Grenoble
Master Informatique
Specialization Artificial Intelligence and the Web

Integration of standards and guidelines into the open source information management systems for biobanks

Nataša Vodopivec

June 24, 2019

Research project performed at Laboratory for computer structures and systems

Under the supervision of:
Miha Moškon

Defended before a jury composed of:
Danielle Ziebelin
Pierre Genevès
Cédric Lauradoux
Georges Quénot

June 2019

Abstract

Recent advances in science and technology enable researchers to discover ever more information in biological, i.e. human, model organism or cell line samples. Finding links between genes, medical data and environmental factors leads to new and more effective ways of diagnosing and treating diseases. The base of every medical research is a large amount of experimental samples, which need appropriate handling and storage to enable high-quality data generation. Collected biosamples are stored in repositories called biobanks. Biobanks are complex systems designed to serve a number of purposes. They must handle the collection of biosamples, enable appropriate storage conditions to safeguard sample integrity, keep track of the samples, enable different access permissions for security reasons and manage informed consents and distribution of biosamples. Furthermore, all processes in the biobank and every step in the management of the biosamples and associated data must be logged. Additionally, biobanks must comply with regulations on protection of personal data and with other standards throughout their procedures. Biobanks are therefore usually supported by powerful information systems, to help with addressing the challenges mentioned above. Biobank Information Management Systems (BIMS) have to cover many details and are therefore not trivial to develop. Many smaller research groups cannot afford to use one of the commercial solutions as these tend to be very expensive. They may also not have funding, time or expertise to develop their own BIMS. In this work, we aim to find an open source BIMS solution that is appropriate to manage a biobank. To be able to compare different open source BIMS solutions, we create a list of minimal requirements that any adequate BIMS should support. In building of the list, we review a selection of essential and commonly used regulations, standards and guidelines and combine them with the knowledge of general functionalities of a biobank. We give special attention to the General Data Protection Regulation (GDPR) of the European Union, General requirements for biobanking ISO 20387:2018, FDA 21 CFR Part 11, ISBER Best Practices and harmonization of data and MIABIS. In each document, we focus on what is important in the context of sample management, which is a focal point of BIMS. Using the list, we evaluate available open source systems by comparing the coverage of requirements of each system. We discuss the results and select the best option. At the end, we comment on what the chosen open source solution is still lacking and how it could be improved.

Acknowledgement

I would like to thank my supervisor Miha Moškon for his invaluable guidance, kindness and comments while reviewing this report. I would also like to thank my advisor Pierre Genevès for all advice he has given me. I would like to express sincere gratitude to my family and friends for always standing by my side and offering me continuous support and encouragement throughout my years of study and through the process of researching and writing this project. I would additionally like to thank AdFutura for enabling me to get my education abroad.

Contents

— 1 —
Introduction

Recent advances in science and technology enable researchers to discover ever more information in biological, i.e. human, model organism or cell line samples. Samples such as blood, urine, tissue and saliva are important resources for medical genetic research that investigates the cause and development of different diseases. Understanding of the disease at the molecular level enables its better detection at the earliest stages and prevention of its development. Medical research aims for prevention, or if that is not possible identification of better treatment options for patients who have already acquired the disease.

Biological samples by themselves are not sufficient sources of information therefore clinical information of patients is often included into the research data. It has been observed that sometimes genes get activated only in combination with certain environmental factors. Several environmental factors, including pesticides, metals, head injuries, lifestyles and dietary habits have been associated with increased disease risk or even with protection [29]. Therefore analyzing large amounts of clinical information taken from patient records can help to identify disease patterns even further. Finding links between genes, medical data and environmental factors leads to new and more effective ways of diagnosing and treating diseases.

Treatments that are very successful for some patients might not work well on others. The patient might have undergone a substantial part or all of the therapy before its efficacy can be assessed. Therapies for diseases like cancer regularly cause serious side effects on the patient. This is undesirable, especially when the treatment has not even been effective. In some cases it is critical to begin with the correct treatment from the start, otherwise the patient's life might be at stake. It is therefore important to understand why this deviation happens and to predict how the patient will respond to a particular treatment. Such knowledge would provide better diagnoses and would improve the selection of appropriate treatment options for patients from the start.

An emerging approach called **precision medicine** [17] is trying to tackle this issue. Known also as **personalized medicine**, it is a new approach to patient care, that aims to make medical care more adapted to each individual. Presently, patients are mostly treated with an one-fits-all general approach that does not help them in equal measures. In precision medicine, patients are to be prescribed targeted treatments that are most likely to help them based on their profiles. This does not mean that the patient will get an individual treatment with medications created uniquely for him or her. Instead, the patients will get classified into subgroups based on the differences in their genes, environmental and lifestyle factors that make them similarly susceptible to certain diseases or treatments. These groups will get tailored treatments that will work

for them better than the general treatment would [18]. Vast amounts of data need to be analyzed in order to find appropriate patient subgroups, so precision medicine is tightly coupled with recent advances in big data technologies. As new classification methods are still being developed, precision medicine will continue to improve as well.

Every medical research needs large amounts of high quality samples. However, vast data pools are difficult to obtain for small research groups as getting samples from a large number of people requires a lot of time, equipment, knowledge and funding, which may not be possible for such groups. Additionally, patient clinical records are frequently needed to complement the biological data. It is highly important that both the data that comes with the biosamples and the clinical records are handled in compliance with the regulations on protection of personal data and with ethical and other guidelines of the European Union. Before any of the samples can be stored or used for research, an informed consent needs to be signed by each participant from which the sample(s) have been taken [15].

In hand with current privacy laws such as General Data Protection Regulation (GDPR) [15], all of the personal data needs to be securely stored so that only the authorized persons can access it. This is because biological samples contain sensitive data about donors such as inclination towards certain diseases that could cause great distress to the donors should this information come into public domain. On top of that there exist methods that allow identification of a person just by their genetic data [13]. These methods are highly technical, albeit possible, so all of the personal data needs to be protected with utmost care [7].

Unless the samples are collected with the intent of being used for a particular research, their analysis does generally not begin right after they have been removed from the body. They need to get stored until they will be needed. Parts of samples also need to be kept for juridical reasons or to be able to prove the results of the research at a later time. It is important that they get stored to a place that will preserve their integrity. Any change of the sample after its collection reduces its quality and might make it unusable for research. The cells and tissues need to be kept structurally intact. Biological samples are usually stabilized by being put to very low temperatures with a process called cryopreservation [26]. They get cooled down to between -80 and -196 degrees Celsius, which is the temperature at which the enzymes cease to operate. This means that chemical processes in the cells stop, so the sample will not change in any way as long as it is kept at low enough temperatures. It is critical that the sample is cooled down with appropriate techniques. This is because the main component of any cell is water – water molecules account for 70% or more of total cell mass [12] – and if the freezing process is not done properly, ice crystals may form and can destroy the sample.

Collected biosamples are stored in repositories called **biobanks**. Biobanks are complex systems designed to serve a number of purposes. They must handle the collection of biosamples, enable appropriate storage conditions to safeguard sample integrity, keep track of the samples, enable different access permissions for security reasons and manage informed consents and distribution of biosamples. Furthermore, all processes in the biobank and every step in the management of the biosamples and associated data must be audited. Additionally, biobanks must comply with regulations on protection of personal data and with other standards throughout their procedures. Biobanks are explained in more detail in section 2.1. It is almost impossible to manage such a system on paper, so information systems are built to take care of it. They are called **Biobank Information Management Systems (BIMS)**. They must be powerful information systems since they have to cover many details and are therefore not trivial to build. Research groups may not have time and funding to build their own BIMS. There

exist many commercial solutions that can help in setting up the biobank, but they tend to be very expensive. Small research teams might not have enough funds to afford such a system. A better option for them is to use an **open source information management system** to run their biobank. However, they want to keep their biobanks competitive with the big commercial biobanks in ensuring the high quality of the biosamples. This is so that the biosamples are trusted to be useful for research. They want to keep the labeling of their biosamples compatible with other biobanks to permit easier exchange of biosamples and information between different research groups, because a bigger pool of biosamples enables better research results. Additionally, open source information systems must comply with the same regulations and standards as the commercial solutions.

In our research, we aim to build a list of minimal BIMS requirements. This list should cover all technicalities and should comply with all of the mandatory and with as many as possible commonly used regulations, standards and guidelines for biobanks. Moreover, we want to find an appropriate open source BIMS that we could use to manage our biobank. We first explain the term biobank in section 2.1. We then list and describe the regulations, standards and guidelines for biobanks in section 2.2. We explain the BIMS in more detail in section 2.3. We build a list of minimal requirements for a BIMS in chapter 3 and introduce the relevant open source solutions in section 4.1. They are compared for their appropriateness in the rest of chapter 4, where the best available open source solution is chosen. We propose how to extend it to be actually able to use it to run a biobank and discuss future work in chapter 5.

— 2 —
State-of-the-Art

2.1 Biobank

The term biobank refers to a biorepository that collects, processes, stores and distributes biospecimen and their associated data [30]. Biospecimen is short for biological specimen and is a sample of biological material or biomaterial. It can be taken from different biological sources such as human, animal, plant, bacteria and others, but we will focus on the human samples such as blood, urine, tissue, cells, feces, DNA, RNA, protein or plasma in this work. A biobank may be a temporary storage for human biomaterial that will be used in clinical procedures and treatments such as blood for transfusion, organs for transplantation, bone marrow for stem cell therapy and others. However, the term most often refers to research biobanks that collect, process, store and distribute biospecimen with the purpose of providing resources for medical studies. They can be located at universities, hospitals, diagnostic laboratories, pharmaceutical laboratories, private organizations or other research institutions. Biobanks that collect human biosamples can be divided into two groups [7]:

- **Population based biobanks** collect biomaterial from all kinds of people, healthy and ill alike, who volunteer to give samples, often complemented with lifestyle data. The samples are used to help identify predisposing factors or susceptibilities to different diseases in the general population and to find groups of people that are more prone to certain diseases. Research projects in population based biobanks can be very long and it may be asked of volunteers to keep giving samples for decades.
- **Disease based biobanks** are integrated into hospitals and collect samples from patients that have already been diagnosed with a certain disease. This is done in order to find out which biomarkers are associated with the disease and how they develop and change during the course of the treatment. A control group of healthy donors is used for comparison.

It may take years to confirm the results of a medical study. Often the donors will not be able to see the immediate results of what has been discovered with their samples and the improved treatments will come too late for them to partake in [7]. However, the studies aim to produce discoveries that will help general population in the future generations.

The **life cycle** of a biospecimen in a biobank comprises of four main stages that are shown in figure 2.1. It is collected, processed, stored and distributed to researchers to be used in studies. The following paragraphs give general explanation of each of the stages.

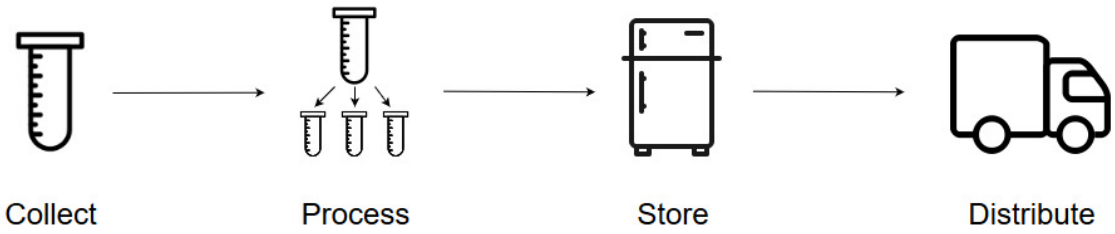

Figure 2.1: Modular structure of a BIMS.

Collection There are different ways in which samples are collected from the participants or patients body. Extra tissue that might otherwise get thrown away can be collected during routine operations as the patient is already there and it does not impose any extra danger for them. Some samples, such as blood or saliva get collected during general checkups, where slightly more of the tissue that would be otherwise is taken. Sometimes the samples are taken with the intent of specific research project and the donors come to medical institutions to give samples with this specific purpose. Lastly, samples that are kept in hospital archives for juridical reasons can be used for research instead of being disposed of when the period in which they needed to be kept has passed. Whether such samples can be re-used or not depends on the regulations of each country. Therefore, different types of samples are obtained more or less easily. If lung samples would be needed for research, collecting them would involve a complicated procedure, which might prove impossible for small research groups that may be lacking the required expertise. Biospecimen need to be collected with existing regulatory standards that ensure high quality of samples. This is generally done by following predefined **standard operating procedures (SOPs)**. Human samples cannot be used and stored without the patient's knowledge or without the approval of an ethical committee. Specimen are generally collected for a specific study. Before the samples can be linked to it, it must have been checked for it's scientific content. The patient is given information about how and for what purpose their sample will be used and for how long it is going to be stored. They are also asked to answer questionnaires on their nutrition, health, lifestyle, smoking and drinking habits, known diseases in the family etc. They must agree if the study wants to use their clinical records. Of all of the collected data, personal data must be treated with the greatest care and follow all of the privacy regulations to protect the patients and donors. During the process of collection is also when the sample custodian is decided – who will own and be responsible for the sample once it has been removed from the body. After collection, biospecimen are brought to the biobank.

Processing Biospecimen are taken to the laboratory where their aliquots and derivatives are produced. Aliquot is a portion of a sample of biological material that has been divided into separate parts [16]. Several samples are made out of a single one so that they can be used in more than one study or analysis. One aliquot generally needs to be kept for juridical reasons or for later validation of the study. Derivative is the processed outcome of parent specimen (e.g plasma or serum from blood) [38]. Aliquots and derivatives should be done by following specific SOPs to ensure continuous high quality of the samples.

Storage The area of storage in a biobank can range from a single fridge to dedicated build-

ings. Biospecimen need to be stored at appropriate temperatures to ensure their quality. Some biological samples can be stored at room temperature, preserved in paraffin blocks or glass slides, but they will keep too little information to be able to be used in studies for which molecular data is desired as the DNA and RNA will degrade too fast. Therefore, biobanks store samples either in mechanical freezers with a temperature around -80 degrees Celsius or in freezers with liquid nitrogen that has temperatures between -196 and 210 degrees Celsius, which is even more suitable for long term storage. The samples should be stored at a relatively constant temperature and should be prevented from thawing as that could lower their quality or make them unusable.

Distribute Collaborations between biobanks allow researchers to get access to data of more people. Bigger sample sizes can lead to easier and faster discoveries of disease patterns. So can samples of a greater amount of genetically and ethnically different people. Moreover, distribution of biospecimen to research groups that do not have their own collections and means to get them allows for more different studies and promotes science. However, not everyone can get the samples. Just as for collecting the samples for a study, the rules remain the same for distributing them. The study for which they have been requested needs to be checked for its scientific content, approved by an ethical committee and safety of the privacy data must be ensured. Only then can the samples be shipped. During the transportation, the samples should not thaw or if they do, this needs to be noted, so usually dedicated companies with specialized containers do the transport between different biobanks. They follow SOPs to preserve quality of biospecimen. Sometimes different biobanks use same labels and annotations for different properties of the biospecimen so confusion may arise. Sharing of data is easier if all of the biobanks follow the same labeling and sample annotation guidelines and thus enable the **harmonization of data**. This is explained in more detail in section 2.2.5. Depending on local regulations, the biospecimen may not be allowed to be kept at the biobank it was sent to. It must either be returned after the research has been done or destroyed with proof.

2.2 Standards and guidelines relevant for biobanking

Due to the nature of the biological samples that are stored and used in biobanks, they have to comply with different regulations, standards and guidelines. In this section we will give comprehensive overview of the most important regulations, standards and guidelines that must or should be followed when setting up a biobank. We have chosen those that appear most often in the recent papers and applications. Additionally, we will give particular attention to the General Data Protection Regulation (GDPR), as it is a law without which the BIMS cannot work properly at least not within the European Union. While reading and explaining the chosen regulations, standards and guidelines we have focused particularly on the parts that must be taken into account when building the list of requirements for a BIMS.

2.2.1 General Data Protection Regulation (GDPR)

The protection of natural persons[1] in relation to the processing of personal data is a fundamental right [15]. This can be concluded from the Article 8(1) of the Charter of Fundamental Rights of the European Union and from the Article 16(1) of the Treaty on the Functioning of the European Union which provide that everyone has the right to the protection of personal data concerning him or her.

General Data Protection Regulation or the Regulation (EU) 2016/679 of the European Parliament and of the Council [15] is a EU-wide legal framework for the protection of personal data [13]. It has been adopted on 27 April 2016 and became binding in its entirety and directly applicable in all Member States of the European Union on 25 May 2018. It is referred to as the "GDPR" or the "Regulation". The main objectives of GDPR are the protection of natural persons with regard to the processing of personal data and the free movement of such data. It states that the free movement of personal data within the European Union shall be neither restricted nor prohibited for reasons connected with the protection of natural persons with regard to the processing of personal data [15].

Here, we overview the sections that must be considered for the lawful working of a biobank. We focus particularly on those sections that are needed when designing a BIMS. They can be roughly divided into seven groups. First and second give introduction to several terms. A group called 'Processing' explains what processing is and what rules and principles it must follow. Next group gives an overview of the rights of the data subjects. The following one details what is a consent. 'Transfer' gives information on how personal data can be transferred outside of European Union. Succeeding group provides how security of the personal data is reached. The last group explains how GDPR is enforced.

Personal data

As defined by the Regulation, **personal data** means *any information relating to an identified or identifiable natural person ('data subject')* [15]. In the context of biobanks, an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data or by reference to one or more factors specific to the physical or genetic identity of that natural person.

Regulation also recognizes **special categories of personal data**. In the context of biobanks the relevant categories are the data revealing racial or ethnic origin, data concerning health, data concerning natural person's sexual orientation and genetic and biometric data when used for the purpose of uniquely identifying a natural person.

Regulation does not apply to the personal data of deceased persons.

Controller and processor

As described in the Regulation, **controller** is a role that means *the natural or legal person, public authority, agency or other body which determines the purposes and means of the processing of personal data*. A **processor** is a role that means *the natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller* [15]. Generally, a biobank serves both purposes and takes both roles.

[1] **Natural person** is a person that is an individual human being, as opposed to a legal person, which may be a private (i.e., business entity or non-governmental organization) or public (i.e., government) organization [32].

GDPR gives controller several responsibilities. Specifically, it must implement appropriate technical and organisational measures and data protection policies to ensure that processing is performed in accordance with the Regulation. It also must be able to demonstrate it's compliance.

Furthermore, the controller must maintain **a record of processing activities** under its responsibility. The requirements of what information must be put in the record are listed in the Article 30(1) and we will not detail them here as they are out of the scope of this section.

Processing

Processing of personal data including the special categories of personal data is prohibited unless the data subject has given explicit **consent** to the processing of those personal data for one or more specific purposes. The requirement of collecting an informed consent from the data subject in biobanks will be explained in detail further on. For now, let us point out that biobanks have consents from data subjects and can thus process special categories of personal data, if this was agreed on in the signed consent. As biobanks process data for scientific research purposes, they can be exempted from this prohibition in accordance with the Article 89(1), insofar as the rights and freedoms of the data subject are safeguarded. For the processing to be **lawful**, the data subject must give a consent as described above.

There are seven principles that need to be followed when processing personal data. In accordance with the Article 89(1), two of them can be exempted for biobanks. They are marked with a '*' in the list below. The principles are:

1. **Lawfulness, fairness and transparency**. Personal data must be processed lawfully, fairly and in a transparent manner in relation to the data subject. Transparency means, that the natural persons should be aware that their personal data are collected, used and otherwise processed. They need to be informed as to what extent their personal data will be processed and about the identity of the controller. Furthermore, transparency means that all information given to natural persons must be easily accessible, easy to understand and that clear and plain language is used.

2. **Purpose limitation***. Personal data must be collected for specified, explicit and legitimate purposes and must not be further processed in a manner that is incompatible with those purposes.

3. **Data minimisation**. Personal data must be relevant and limited to what is necessary in relation to the purposes for which they are processed.

4. **Accuracy**. Personal data must be accurate and where necessary, kept up to date.

5. **Storage limitation***. Personal data must be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed.

6. **Integrity and confidentiality**. Personal data must be processed in a manner that ensures appropriate security of the data. This includes protection against unauthorised and unlawful processing and against accidental loss, destruction or damage. It should be done using appropriate technical and organisational measures.

7. **Accountability**. The controller of the data shall be responsible for compliance with the previous six principles and must be able to demonstrate this compliance.

Rights of the data subject

A data subject is given a set of rights, which includes a right to access, to rectification, to erasure, to restriction of processing, to data portability and to object.

Right to access by the data subject is the right to get confirmation about whether his or her personal data is being processed in any way. If yes, he or she should be given information about processing as detailed in the Article 15(1). **Right to rectification** permits data subject to complete incomplete or inaccurate personal data. **Right to erasure**, more known as **right to be forgotten** is one of the most important additions to the rights of a natural person by the Regulation. It gives controller the obligation to erase all personal data of a subject, should they wish so (for example by withdrawing consent) or if other conditions for erasure apply. Following this right, all personal data must be deleted as soon as they are no longer necessary in relation to the purposes for which they were collected. **Right to restriction of processing** allows data subject to restrict processing of his or her personal data. **Right to data portability** is the right of a data subject to receive personal data from a controller. It should be provided in a structured, commonly used and machine-readable format so that it can be transmitted to some other controller. It applies to biobanks because the processing there is based on a consent. **Right to object** permits the data subject to object to processing of personal data concerning him or her.

When data subject invokes any of above the rights, the controller has the obligation to provide him or her information on what action has been taken. The controller also has the possibility to inform the data subject that no action will be taken and in that case data subject may lodge a complaint. Controller has to react within one month of receipt of the request and may request that the data subject confirms his or her identity.

Aside from the rights of data subjects listed above, biobanks collecting personal data have another obligation. Under the Regulation, they must provide their participants with extensive information about what personal data is processed and how. This has to be done at the time when data is obtained *and* when it is updated. Since personal data can be collected directly from a data subject or acquired from third parties, different information must be provided to the data subject depending on this case. What exact information must be given is defined in Article 13 and Article 14 of this Regulation and we will not detail it here as it is out of the scope of this chapter.

However, GDPR allows for **exceptions to various data subjects' rights**. A number of these exemptions may apply directly on a case-by-case basis while others will have to be provided by Union or Member State law [13]. Since biobanks collect data for research purposes, they fall under one of the specific processing situations, namely the Article 89. It says that Union or Member State laws can provide for exceptions from the right of access, right to rectification, right to restriction of processing and the right to object. Derogations are allowed only when the aforementioned rights are likely to seriously impair the achievement of objectives of specific purposes or to render it impossible. Because of Article 89, the right to be forgotten also does not need to apply if it is likely to affect the results of processing. It is important to note that derogations can only be done under appropriate safeguards. They have to ensure that technical and organisational measurements are in place, in particular to ensure the respect for the principle of data minimisation (e.g. by pseudonymisation).

Consent

By GDPR the processing of personal data must be lawful. One way to meet this requirement is by a **consent**, which is a common practice in biobanks. Consents for each project are generally stored within a BIMS. A consent is a freely given, specific, informed and unambiguous indication of the data subject's agreement to the processing of personal data relating to him or her. Specific means, that the data subject should be aware of the fact that he or she is giving a consent and what it covers. It should cover all processing activities that will be carried out. Silence, pre-ticked boxes or inactivity therefore do not constitute consent. Freely given means that the consent should not be forced out of a data subject in a way that the provision of a service is conditional on consent to the processing of personal data that is not needed for the functioning of the service. For consent to be informed, the data subject should be aware at least of the identity of the controller and the purposes of the processing for which the personal data are intended. Finally, the term unambiguous refers to the principle of transparency. It states that the consent should be given to the data subjects in an intelligible and easily accessible form and should use clear and plain language. If data subject is a child, a language that him or her will understand must be used. For biobanks, consents are usually collected in a written or electronic form.

There are other conditions to consent under the Regulation. Firstly, the data subjects must be informed that they have a right to withdraw their consent at any time. The process of withdrawing a consent must be as easy as giving one. Secondly, children merit a specific protection with regard to their personal data as they may be less aware of the risks, consequences and safeguards in relation to the processing of personal data [13]. Unless a Member State provides a law for a lower age, the child must be above 16 years old in order for the processing of personal data to be lawful. For younger children, the consent is given by the holder of parental responsibility over a child. Thirdly, the controller should be able to demonstrate that the data subject has given consent to the processing operation.

However, GDPR acknowledges that in the domain of scientific research it is often not possible to fully identify the purpose of personal data processing at the time of data collection. It therefore allows that data subjects give their consents to certain areas of scientific research, as long as they are within the recognised ethical standards. This could be referred to as a '**broad consent**'. It also allows data subjects to give consent only to selected areas of scientific research or to parts of research projects insofar as these partial consents are still sufficient for the research to reach its objective.

Transfer

Biological samples and their associate data can be sent from one biobank to another for the research purposes. Free movement of personal data is one of the main objectives of GDPR. It defines rules for transfers outside the European Union. Any transfer of personal data which are undergoing processing or are intended for processing after transfer to a third country or an international organisation must ensure that the level of protection of natural persons guaranteed by GDPR is not undermined. Regulation defines three cases where the personal data can be transferred outside of European Union [13]. Firstly, when the European Commission has decided that the third country ensures an adequate level of protection (called 'adequacy decision'). Secondly, in the absence of adequacy decision the controller and processor in the

third country must provide appropriate safeguards and have enforceable data subject rights and effective legal remedies. Thirdly, if none of the above apply, the transfer may still take place in the data subject explicitly consents to if after having been informed of the possible risks. We will again not detail these concepts any further as they are not intrinsic parts of BIMS.

Security of personal data

Biobanks process a lot of personal data and since protection of natural persons in relation to the processing of personal data is a fundamental right, it is imperative to take appropriate precautions for safeguarding it and to penalize infringements.

Technical safeguarding measures emphasized in the Regulation are the pseudonymisation and encryption of personal data. **Pseudonymisation** means *the processing of personal data in such a manner that it can no longer be attributed to a specific data subject without the use of additional information* [15]. Such additional information should be kept separately and be subject to measures that ensure the personal data are not attributed to an identified or identifiable natural person. **Anonymised data** means the *information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable* [15]. The Regulation does not concern the processing of anonymous information.

Alongside the techniques described above, Regulation advises additional organisational measures to help ensure an appropriate level of security. These are [15]:

- The ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services;
- The ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident;
- A process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing.

There are many possible risks to the rights and freedoms of natural persons in case the personal data leaks out to unauthorised persons. Where such personal data are processed which reveal racial or ethnic origin, genetic data or data concerning health, risks to the natural person can be inter alia damage to reputation, discrimination and economic or social disadvantage. So the controllers (such as biobanks) processing these data should know that the safety of it should not be taken lightly. By the Regulation every biobank must make a **data protection impact assessment** prior to processing personal data to see how high a risk the loss of it would be. Additionally, a **data protection officer** needs to be appointed, that will assist in monitoring internal compliance with the GDPR. A biobank should be able to obtain a **certification** that demonstrates its compliance with GDPR. In case of a **personal data breach** the supervisory authority and the data subject should be notified about it without delay. We will not detail these concepts as they are not intrinsic parts of BIMS.

Enforcement of the Regulation

The Regulation provides for three types of mechanisms to enforce its provisions: corrective measures, fines and penalties [13]. Within each Member State there exists a public supervisory

authority that gives a set of **corrective measures** which include issuing warnings or reprimands, imposing a limitation or a ban on processing, ordering the rectification or erasure of personal data and imposing an administrative fine to the controller or the processor. **Administrative fine** of up to €20.000.000 can be given to infringement of the basic principles for processing, including conditions for consent. These fines also apply to infringements of data subjects' rights in transfers of personal data to a recipient in a third country or an international organisation. Other **penalties** applicable to infringements of this Regulation exist, in particular for infringements which are not subject to administrative fines. The rules for such penalties are laid down by each Member State.

2.2.2 General requirements for biobanking ISO 20387:2018

The International Organisation for Standardization (ISO) is a worldwide federation of national standard bodies (member bodies) [25]. The work of preparing International Standards is normally carried out through ISO technical committees. Technical committee ISO/TC276 standardizes in the field of biotechnology processes that include biobanking among its topics. ISO/TC276 published a standard called **General requirements for biobanking** or ISO 20387:2018 in August 2018 [25].

This standard was developed with the objective of promoting confidence in biobanking [25]. It applies to biobanking of any biological material (biomaterial) e.g from multi-cellular organisms such as human and animal; from fungus, plants and microorganisms. It contains requirements that will enable biobanks to demonstrate competent biobank operation and quality control requirements that will give biobanks the ability to provide biological material and its associated data of appropriate quality for research and development. This is intended to be achieved first by planning and then by implementation of policies, processes and procedures covering the life cycle of biological materials and their associated data. Furthermore, the use of this document facilitates cooperation, fosters exchange and assists in the harmonization of practices among biobanks, researchers and other parties [25].

The standard consists of eight parts and two annexes. First three parts give the scope of this standard, normative references, terms and definitions and we do not detail them per se. Any unknown terms are explained when needed. Fourth part gives **general requirements** for the operation of a biobank. It emphasizes that biobanks need to be committed to **impartiality**, shall not allow any internal and/or external pressures on it and shall eliminate them should they arise. It also emphasizes **confidentiality** of the biobank which we detail later on together with part five, which gives **structural requirements**. Here, we have noticed many similarities with GDPR and we point them out. The next part gives **resource requirements**. Biobank's resources are the personnel, facilities/dedicated areas, equipment, information system and support services [25]. This part first describes how personnel needs to be trained, tested on their competence and how they should perform their duties. It then provides rules on how the facilities, dedicated areas and equipment need to be set up and be taken care of. Lastly, it gives requirements for externally provided processes, products and services. As these apply to the internal working of a whole biobank and not to the BIMS, we do not discuss them further. Seventh part lists **process requirements**. We have focused a lot of our attention on this part. The last part gives requirements for the **quality management system (QMS)**. It lists how a biobank should document and manage the information necessary for its planning and operation in order to comply with this ISO standard. It states that a biobank should be capable of

supporting and demonstrating the consistent achievement of requirements that are presented in this standard. As the requirements in this part do not include direct requirements for the information management system, we do not discuss them further.

The first annex gives the documentation requirements relevant for the biological material and the second annex provides complimentary information to help implement them. This part of the standard promotes the **interoperability** of a biobank, which is a very important notion as it facilitates exchange of information and samples between different biobanks. We will expose the aforementioned requirements later on in section 2.2.5 where we will compare them with the popular Minimum Information About BIobank data Sharing (MIABIS) guidelines [28].

As stated in the above paragraph not all parts of this ISO standard apply for the BIMS. Many define the rules on how to run a biobank as a whole system and on how to make it competent. Here, we overview parts that provide useful information for BIMS. The confidentiality and structural requirements are covered together and the process requirements separately from them. We have introduced all of them in the above paragraph. Additionally, ISO defines a set of information about the biobank that must be always clearly written down:

- The biobank's mission.
- The information relevant to biobank activities, processes and procedures.
- The identities of personnel that are performing any activities on the biological material.
- The identity or identities of top management that has overall responsibility for the biobank.
- The document with the range of activities for which the biobank conforms with this ISO.

Confidentiality and structural requirements

The standard states that the biobank should be legally responsible for all its activities. This means that it should be responsible for management of confidential information through legally enforceable commitments. This complies with the GDPR that provides mechanisms to enforce its provisions as explained in section 2.2.1. Furthermore, the biobank shall protect confidential information and proprietary rights of donors. GDPR provides a good way to do that through pseudonymisation which is defined in section 2.2.1.

The biobank should identify top management that has overall responsibility for the biobank but in addition to that an advisory board should be appointed that will guide and advise management on scientific, technical and/or administrative and other matters [25]. A data protection officer that was presented in the section 2.2.1 could be a part of this advisory board to comply with the GDPR. All of the personnel having access to data should be bound to confidentiality, to protect the personal data of the donors, which is an extension of the GDPR. Atop of all that, the biobank should have a course of action to define and address liabilities arising from its activities, which could be covered by the data protection impact assessment from section 2.2.1, again to comply with the GDPR.

The information regarding biological material and associated data can only be released according to relevant agreements and approvals (e.g. contractual agreements, legally binding documents, ethical approvals) [25]. When sharing the data or biological material, the biobank should inform the donor on how their privacy is protected (where possible).

Process requirements

This standard states that a biobank must establish, document and implement **standard operating procedures (SOP)** addressing all stages of a life cycle and all other processes relating to biological material and associated data. In this section we first explain what every SOP must include in the general context of a biobank and then we list all of the stages of the life cycle, define them and provide additional information for each of them that is relevant for the BIMS. We have already explained the four main stages of the life cycle in section 2.1. Here, the life cycle is presented in more detail and thus comprises of more stages.

All procedures should be documented with regard to the specific biological material. All **critical activities** within each procedure should be identified and documented, where a critical activity is an activity that *might have an impact on the fitness for the intended purpose of biological material and/or associated data* [25]. A SOP for **quality control** (QC) must exist for and be used at each identified critical activity. Quality control should also be performed at planned intervals of regular frequency to ensure the non-changing quality of the material. Biobank must retain documented information of QC activities and the results. If a predefined quality criteria is not met, actions should be taken to control reporting of invalid data and/or distribution of non-compliant biological material. The methods that are used for critical activities must first be **validated** in order to ensure fitness or the intended purpose and then **verified** by the biobank. The procedures used for validation and verification must be documented. Furthermore, all of the procedures should ensure compliance with relevant bio security and bio safety requirements and address risks and opportunities using a risk assessment [25].

Apart from the SOPs for the life cycle processes the standard states that SOPs for three other processes must also exist. First, a SOP to support the patient/donor right to **consent withdrawal** for storage and use of their biological material and associated data. Second, a procedure that defines how to make and what to include in a **report** of the current state of the material in the biobank. Third, procedures on how to receive, evaluate and make decisions on **complaints**.

All SOPs should be available to personnel.

The life cycle of a biological material consists of collecting, acquiring/receiving, tagging, storing, preparing, preserving, cataloging, destroying, distributing, packaging and transporting. The BIMS should provide an option to mark which SOP was used at each stage and an option to flag any deviation. We systematically list all of the stages of the life cycle and define them where needed. We provide what additional information, if any, is relevant for the BIMS for each of the stages:

- **Collection**. When the sample is collected at least the information about the date, time, place and procedure of the collection process must be noted. The interoperability of this information should be supported. If the donor of the sample is human, an ethical approval must be acquired before the sample can be taken.

- **Acquisition or reception** is *the act of obtaining possession or custody of biological material and/or associated data* [25]. This means that the biobank was not responsible for the collection. The SOP for the acquisition must include acceptance criteria for the biological material. The same minimum information about the sample must be noted at the time of the acquisition as in the case of collection, but there should be an option to note how the sample came into the biobank, to differentiate between the two options. If

needed by a particular biobank and if possible, the information related to stages prior to the reception should be documented.

- **Tagging**. Biological material must be tagged with **unique identifiers** and use persistent tagging that can be maintained through the whole life cycle under the custody of the biobank [25]. This introduces the term **traceability**. It states that it must be possible to identify the location of the biological material and associated data at all times. It further states, that there must exist a link between biological material and its associated data. Additionally, each sample should have a link to the information with details on permissions or restrictions associated with its use. Note that this last requirement is similar to the requirement of a consent under the GDPR.

- **Storage**. For any sample in storage two things must be registered. Firstly, the tagging information including at least the unique identifier of a biological material and secondly the information about the type of container and environmental conditions of the storage where it is. Whenever a sample is moved or accessed the date, time and the personnel having done it must be noted.

- **Preparation** by the definition are *the activities taking place in a laboratory after acquisitioning to make biological material ready for further use in the life cycle* [25]. These activities are centrifuging, purifying, fixing stabilizing, aliquoting and others. We're taking the assumption that the samples arrive to the BIMS in their prepared state so we do not discuss this stage further.

- **Preservation** is *the act to prevent or retard biological or physical deterioration of biological material* [25]. During the process of preservation any changes should be monitored and the material disposed if its quality gets ruined.

- **Cataloging**. The biobank should provide interested parties access to a catalog of available biological material and associated data.

- **Disposal**. Records of biological material and associated data that have been disposed should be kept.

- **Distribution**. Before any distribution a documented agreement outlining the conditions on how biological material and the associated data can be used must be signed. A report of what is included in the package must be provided.

- **Packaging and transport**. The SOP for these stages must include procedures for safe handling and packaging of both the biological material and the data. Prior to shipping, a documented agreement must be signed. The critical chain-of-custody (who is handling the package) and the elements that may compromise the quality od the biosample must be maintained from the point of dispatch to the point of receipt.

2.2.3 FDA 21 CFR Part 11

Title 21 CFR Part 11 [9] is the part of Title 21 of the Code of Federal Regulations (CFR) that establishes the United States Food and Drug Administration (FDA) regulations on electronic records and electronic signatures (ERES) [10]. Part 11, as it is commonly called, defines the

criteria under which electronic records and electronic signatures are considered trustworthy, reliable, and equivalent to paper records. These requirements for the electronic records include [9]:

- Validation of systems to ensure accuracy, reliability, consistent intended performance, and the ability to discern invalid or altered records.
- Protection of records throughout their retention period and the the ability to generate accurate and complete copies of records.
- Procedures should be in place to limit system system access to authorized users. Limited access can be ensured through physical and/or logical security mechanisms. For logical security users typically have to log on to a system with a user I.D. and password. Physical security through key locks or pass cards can serve as an additional safety measure [11].
- Use of secure, computer-generated, time-stamped audit trails. On any action (user log-in/log-out, user action, moving of the biosample, user note, creation, modification or deletion of data, etc.) the computer should record what action was made, who made it and its date and time. This is to ensure data integrity.
- Procedures should be available to use operational system checks to enforce permitted sequencing of steps and events, as appropriate.
- Use of authority checks to ensure that only authorized individuals can use certain functions of the system. The system should be able to verify that an individual is permitted or authorized to perform the requested function.
- Use of device (e.g., terminal) checks to determine, as appropriate, the validity of the source of data input or operational instruction.
- The persons who use the electronic records have appropriate education or training.
- The establishment of written policies that hold individuals accountable and responsible for actions initiated under their electronic signatures, in order to deter record and signature falsification.
- The access to the system documentation should be controlled.

Electronic signatures which meet the requirements of the rule are considered to be equivalent to full handwritten signatures. The requirements for handwritten signatures include [9]:

- Each electronic signature shall be unique to one individual and shall not be reused by or reassigned to anyone else.
- The organization needs to verify the identity of the individual before it can establish, assign, certify, or otherwise sanction his or her electronic signature.
- Prior to or at the time of use of electronic signatures, the organizations need to certify to the agency that the electronic signatures in their system are intended to be the legally binding equivalent of traditional handwritten signatures.

The requirements that come particularly into account for building a BIMS are: the validation of the system, protection of records and generating their copies, role-privileged limited access, audit trail, procedures for operational system checks, controlled access to system documentation and verified and unique handwritten signatures. Broad sections of the regulation have been challenged as very expensive and for some applications almost impractical [11] and the FDA has stated in their guidance that it will exercise enforcement discretion on many parts of the rule. In practice, the requirements on access controls are the only part routinely enforced.

2.2.4 ISBER Best Practices

International Society for Biological and Environmental Repositories (ISBER) [21] is a global biobanking organization that addresses harmonization of scientific, technical, legal, and ethical issues relevant to repositories of biological and environmental specimens. It is also one of the organizations that makes effective contributions to the work of the technical committee ISO/TC276 which produced the ISO 20387:2018 for biobanking explained in section 2.2.2.

It consists of different working groups, one of which is the informatics working group [22]. In 2013 this group has published an **information system evaluation checklist** of effective practices for the management of specimen collections and repositories [23]. In the list, a selection of practices with the notation **Best Practice** marks the most effective practices. Many of them are already covered by the GDPR, ISO 20387:2018, other articles or are common sense so we will not discuss them here specifically. We included all of the best practices that we thought sensible into our list of minimal requirements for BIMS in chapter 3.

2.2.5 Harmonization and MIABIS

Some research goals would not be feasible without a large biobank collection. **Harmonization** of data and samples is one of the critical enablers for facilitating the collaboration between biobanks. It enables a greater effective utilization of biobanks [19]. Harmonization of biosamples is achieved by following the same SOPs and the harmonization of data is achieved by collecting the same information and by using common ontologies (such as medical ontologies SNOMED [42] and ICD [20]).

One of the first guidelines to tackle this issue was Minimum Information About BIobank data Sharing (MIABIS). It was developed in 2012 by the Biobanking and BioMolecular Resources Research Infrastructure of Sweden (BBMRI.se). After wide acceptance of first version of MIABIS, a working group was formed under the largest infrastructure for health in Europe, Biobanking and BioMolecular Resources Research Infrastructure (BBMRI-ERIC) to continue the development of the guidelines [28]. The current MIABIS 2.0 Core [28] offers three main components. **Biobank** represents an organization or an organizational unit that stores samples and data related to the samples. **Sample Collection** represents a set of samples with at least one common characteristic. Lastly, **Study** represents a set of samples brought together in the context of a research study. They are described by 22 attributes. Certain attributes such as *Contact Information* and *Disease* are described as structured data and are reused by several components. Whenever appropriate, the attribute values follow well known medical terminologies such as ICD [20] and SNOMED [42]. The attributes of data describing Biobank are shown in appendix in table A.1, the attributes of data describing Sample Collection in appendix in table A.2 and the attributes of data describing Study in appendix in table A.3. Five additional components

describing the concepts of Biological Experiment, Sample Donor, Rare diseases, Samples and Quality [28] have been suggested by the MIABIS working group. They are still in development so we have not included them here.

By ISO 20378:2018 the biobank mission and the top management of the biobank should be available. Both are covered by the attributes of MIABIS 2.0 Core. Biobank mission is the same as the *Description* attribute and the top management is the same as the *Juristic person* attribute of Biobank component.

The components of MIABIS 2.0 Core represent high level information of a biobank and biomedical research. For increased harmonization of a biobank, they can be combined with the documentation requirements for biological material and associated data from the annex of ISO 20378:2018. It provides documentation requirements for Acquisition, Storage, Preparation/preservation, Testing, Distribution and disposal and Transport of biological material and associated data. The Preparation/preservation and Testing are not a part of a BIMS but part of a LIMS. All other documentation requirements should be used in the BIMS and are presented in appendix in tables A.4, A.5, A.6 and A.7. Once the MIABIS working group finalizes any of the suggested components these should be added to the BIMS. If any attributes will overlap with those of ISO 20378:2018, an appropriate combination of both should be used.

2.3 Biobank Information Management System

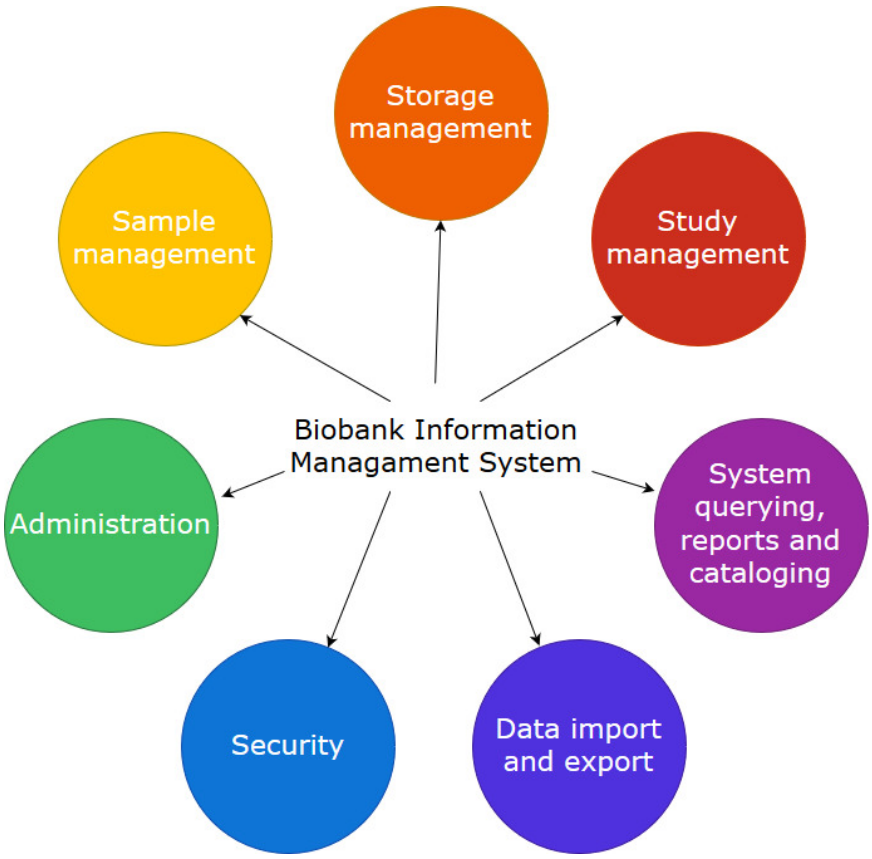

Figure 2.2: Modular structure of a BIMS.

A Biobank Information Management System (BIMS) is an important building block of any biobank. It is a software system that facilitates management of a biobank. Its main functionality is sample management. It does not manage the execution of laboratory analyses and processes (i.e. molecular genetic analyses, extraction of DNA or RNA) on biological material and is therefore not a Laboratory Information Management System (LIMS). It assumes that at the time of its arrival to the biobank, the biological material is already in the form (aliquot, derivative, parent sample) in which it is meant to be stored. Main objects of BIMS are biological material and its associated data and not patients or donors.

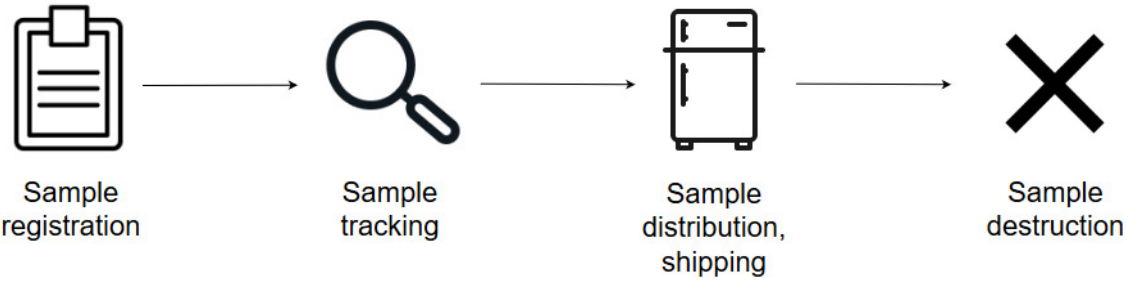

Figure 2.3: Steps of sample management.

BIMS consists of multiple modules that each group related services. The modular structure of BIMS is shown in figure 2.2. It consists of the following modules:

- sample management,
- study and project management,
- storage management,
- system querying, reporting and cataloging,
- data import and export,
- security,
- administration.

The module on sample management further consists of the following modules:

- sample registration,
- sample tracking and management,
- sample distribution and shipping,
- sample destruction.

They are shown graphically in figure 2.3. Each of the modules is first described in more detail and then its minimal functionality requirements are listed in chapter 3.

— 3 —
Minimal requirements for a BIMS

Although technical capabilities in biobanks are low, it cannot be expected that users will simply install and use the software once it becomes available. Ongoing evaluation and communication concerning the requirements will be a continuing effort [31]. This is why we make a list of minimal requirements that will always need to be a part of BIMS, even as the field of biobanking will continue to evolve. There already exist some sets of requirements for a BIMS, such as [14] and [6], but we have taken them a step further by extending them with expectations of an array of research articles and applying the new GDPR [15] and ISO 20378:2018 [25] standards to them. We have not used all of the requirements from the previously mentioned resources as we needed a list of requirements that can be used to compare different open source systems. As such, it needed to be of appropriate length. This is another reason why we have decided to build a list of minimal requirements instead of a list of all possible requirements.

In the following paragraph we first give a rough description of requirements for each of the modules of the BIMS that were introduced in section 2.3. We then list the requirements for each module. This produces the list of minimal requirements for a BIMS which we will use in the following chapters to compare and choose the most appropriate open source BIMS.

1. **Sample management**

   1.1. **Sample registration**

   Misplaced or badly stored samples may cause unwanted delays and increased costs [27], so they should be marked with globally unique IDs. They need to be marked with barcodes to enable their traceability.

   Personal data of donors needs to be encoded. Pseudonymisation is often more appropriate to be used in biobanks than anonymisation. This is because anonymisation of personal data means, that it is de-identified which in turn makes it impossible to connect to the donor. We argue that anonymisation is not a good thing for biobanks out of two perspectives. Firstly, if genetic data is stored in a biobank, it is not possible to anonymise it perfectly. Specific methods allow to always find a person by using their genetic imprint. They are expensive and used mostly in criminal investigation, but it can be done nevertheless. Secondly, being able to track the biosample back to the donor is usually a wish of the biobank. In this way, donor may be asked to give additional samples. He or she can also be told results of the research or in case some disease is unexpectedly found in their samples, they can be made aware of it. Furthermore, anonymisation deprives the donor of his or her

right to withdraw a consent and to be forgotten. Therefore, our BIMS would aim for pseudonymisation. It can be safely done with the use of a good one way hash function on the personal identification number of a person. If the key is stored apart from the original data, this is a relatively secure way of encoding personal data.

To allow harmonization of the data, sample registration form should be as defined by the ISO 20387:2018. But since different biobanks have different sample information needs, this form should be customizable and allow extensions with extra fields. A predefined SOP is chosen to ensure the sample quality, but should have an option to flag and describe deviations of the process. Sometimes, the samples are not collected by a biobank but acquired, so scanning of their existing barcodes should be enabled. It should be possible to add samples in bulk in order to save time. It should be possible to link aliquots and derivatives to their parent samples. The system should allow to note problems in the collection process. Samples should have an option to be linked to a study for which they were collected. If they are collected for a study, the study should be entered into the system first. This is to ensure that the ethical approval is in order. It is also needed because users of the system have limited access to certain studies and they can only see the data of those studies. A informed consent must be linked or uploaded at the time of sample registration. It should be noted whether the consent has already been signed or not.
The system should provide the functionality to:

1.1.1. Enforce globally unique IDs for each sample entry.
1.1.2. Enforce globally unique IDs for each donor.
1.1.3. Generate and print barcodes, RFID and/or other machine readable labels.
1.1.4. Be able to encode the personal data of human donors.
1.1.5. Sample registration form should use documentation requirements from ISO 20387:2018.
1.1.6. Extend the sample registration form with customizable fields.
1.1.7. Have an option to mark whether the sample has been collected or acquired.
1.1.8. Allow to choose a SOP that was used for collection/acquisition.
1.1.9. Allow to flag and describe the deviation of the SOP for collection or acquisition process.
1.1.10. Import samples by scanning their existing barcodes.
1.1.11. Bulk import samples (e.g. whole boxes or plates or trays).
1.1.12. Allow linkage of aliquots to the parent sample.
1.1.13. Allow linkage of derived samples to the parent sample.
1.1.14. Allow to add notes on possible problems of the sample (bad quality for research, physical condition, leaks).
1.1.15. Add links to patient files.
1.1.16. Link the sample to a study.
1.1.17. Link to or upload a signed informed consent form (for human subjects).
1.1.18. Add information about the status of informed consent of a sample (signed, not-signed).
1.1.19. Validate the forms before accepting them.

1.2. **Sample tracking**

All actions performed on a sample and all of its movements inside of the biobank must be audited to comply with the regulations and to safeguard the integrity and quality of samples. The changes that need to be audited are i.e. sample's location, temperature at that location, its freezing and thawing cycles. The audit trail must include at least the date, time and the contact of the personnel that accessed the sample. Users should have an option to add notes on samples. The status of informed consent of a sample should be modifiable (e.g. full or partial withdrawal of consent by the donor, right to object and right to restriction of processing). When the right to be forgotten is invoked and approved, the system should have a procedure to delete all data from a certain donor. The records of removed samples and their associated data should be stored for a predefined amount of time, after it has been de-identified.

The system should:

1.2.1. Have a secure, computer-generated, time-stamped audit trail of all changes/activities associated to a sample (from receiving a sample until distribution/destruction). It must include the date, time and who accessed the sample.

1.2.2. Allow users to add additional information/notes of events (e.g. temperature logs) to specific samples.

1.2.3. Allow users to change information about the status of informed consent of a sample (e.g. full or partial withdrawal of consent by the donor).

1.2.4. Have the procedure to delete all data from a certain donor.

1.2.5. Anonymise the records of checked-out samples and store them for a predefined amount of time.

1.3. **Sample distribution and shipping**

The user should be able to create a list of samples that they would like to take from the freezers either by selecting each sample by itself or by uploading a pre-made list of samples. The system should have a functionality to process a sample request, together with generating and storing appropriate transfer agreements which safeguard the sensitive data of the samples. The procedure of first distribution and then shipping should follow the predefined SOPs. Information requirements should be that of the ISO 20387:2018. The chain-of-custody must be maintained throughout the whole procedure of distribution and shipping. There should be an option to note the elements that may compromise the quality of the sample while in transfer. A note of delivery should be collected after successful distribution. Records of distributed samples should be stored in the system for a predefined amount of time.

The system should:

1.3.1. Allow the user to create a list of samples for distribution

1.3.2. Allow the user to generate a list by uploading a list of samples.

1.3.3. Have the functionality to process a sample request.

1.3.4. Follow the SOP for distribution and shipping.

1.3.5. Have the distribution form from the documentation requirements of ISO 20387:2018.

1.3.6. Have the functionality to generate or store sample transfer agreements.

1.3.7. Have the transfer form from the documentation requirements of ISO 20387:2018.

1.3.8. Maintain the chain-of-custody.

1.3.9. Enable to note the elements that may compromise the quality of the sample while in transfer.

1.3.10. Note the proof of delivery when sample gets distributed.

1.3.11. Store records of distributed samples for a predefined amount of time.

1.4. **Sample destruction**

There are many reasons for the destruction of a sample. It might be for example because the donor has requested it, because the sample has been damaged or has lost its quality or for the lack of space in the biobank. The system must enable complete removal of a subject and all of his samples/instances and other information from the database. The process of destruction must follow a predefined SOP. The proof of destruction must be provided and the records of destructed samples including the reason for destruction must be stored for a predefined amount of time.
The system should provide the functionality to:

1.4.1. Allow complete removal of a subject and all of his samples and other information from the database.

1.4.2. Have a SOP for destruction of a sample and associated data.

1.4.3. Note the proof of destruction when the sample gets destructed.

1.4.4. Store records of destructed samples including the reason for destruction for a predefined amount of time.

2. **Study management**

To enable harmonization, a new study should be entered into the system using the form defined by MIABIS as seen in section 2.2.5. Users are added to the study to enable them to see the data of this study and limit their access to other information of the biobank. The SOPs of the processes of the study should generally be the same as those set up by the biobank to enable interoperability. If some are different, it should be possible to add them to the system. The system should also store ethical approvals and informed consents of the study. There should be an option to see the list of all samples used in the study.
The system should provide the functionality to:

2.1. Add a new study using the MIABIS form for Data describing Study.

2.2. Add users to the study.

2.3. Store the SOPs of the study if they differ from the biobank's.

2.4. Store ethical approvals of the study.

2.5. Store informed consents of the study.

2.6. Show samples used in the study.

3. **Storage management**

Storage facilities in biobanks consist of different types of storage containers within which are hierarchies of smaller containers. An administrator of the biobank and other personnel with appropriate permissions should be able to set up a hierarchy of storage containers that are particular for the biobank. The containers might change structures, so this

functionality in the system should be flexible to permit changes. Storage of the samples should follow predefined SOPs to put them in the correct storage places, based on their types. The information requirements should be the same as in the ISO 20387:2018, described in section 2.2.5. There should be an option to move samples within the existing storage places. The storage conditions should be monitored at all times.
The system should provide the functionality to:

3.1. Create and configure different types of storage containers (tanks, freezers, shelves, trays, boxes).

3.2. Create and configure a hierarchy of storage containers

3.3. Create and configure different types of boxes/trays/samples in a freezer.

3.4. Have a SOP for storage of different types of samples.

3.5. The form should follow the form from the documentation requirements of ISO 20387:2018.

3.6. Reorganize/move samples in existing freezers.

3.7. Monitor container conditions (e.g. deviations of temperature, humidity).

4. **System querying, reports and cataloging**

The distribution of samples and associated data requires sufficient information capabilities such as catalogs to deliver data to external researchers, interested in using the biobank's samples. The catalog should include only non-sensitive data. To find specific information in the system, the system should support querying of the biobank. To ease the searching, some predefined queries should be available. To get more particular information, custom queries with the search by any parameter or their combination should be possible. The user should have an option to save a query for future reuse. Users should be able to find specific stored samples by using their barcodes or IDs. System should provide the ability to generate both predefined and customized descriptive reports and summaries of what is stored in the system.
The system should:

4.1. Allow end-users to search/query the biobank by using predefined queries.

4.2. Allow end-users to search/query the biobank by using custom queries.

4.3. Provide ability to save queries for future reuse.

4.4. Allow end-users to find stored samples by barcodes or by IDs.

4.5. Allow external researchers to search a sorted catalog of what is stored in the system.

4.6. Allow end-users to generate predefined descriptive reports and summaries of what is stored in the system.

4.7. Allow end-users to generate customized descriptive reports and summaries of what is stored in the system and to save them.

4.8. Generate periodic reports of the system.

5. **Data import and export**

To facilitate interoperability, the system should allow to import legacy data (e.g. from other BIMS systems and excel files). It should allow to export data to multiple file formats (e.g. csv, tsv). This is mostly so that it can be imported to other BIMS. The user should be able to export only the data for which they have access permissions.
The system should:

5.1. Allow end-users to import legacy data (e.g. from other BIMS systems and excel files).

5.2. Allow end-users to export data (e.g. to csv files and to other formats).

6. **Security**

The system should provide many security measures. Some to safeguard the sensitive information by securing the access to the system by firewalls and and others to safeguard the integrity and quality of the data. As already mentioned in the section about sample management, the system should keep the personal information of donors separately from the coding key. It should ensure secure storage, usage and follow up of the information. The system should use the secure, computer-generated, time-stamped audit trails. The computer should record what action was made, who made it and its date and time on any action (user log-in/log-out, user action, moving of the biosample, user note, creation, modification or deletion of data, etc.). In the case of power outages or other failures of the system, it should have regular backing-up of all data and the ability to restore the availability and access to personal data in a timely manner. Furthermore, the system should be audited to receive accreditation, integrated in local authentication procedures, should allow only authenticated access and support secure connection protocols (e.g. SSL certificates) [14].
The system should:

6.1. Keep the personal information of donors separately from the coding key and ensure the secure storage, usage and follow up of the information.

6.2. Support the use of secure, computer-generated, time-stamped audit trails. The computer should record what action was made, who made it and its date and time.

6.3. Should have regular backing-up of all data and the ability to restore the availability and access to personal data in the event of a physical or technical incident.

6.4. Should be audited [14].

6.5. Should allow only authenticated access [14].

6.6. Should support secure connection protocols (e.g. SSL certificates) [14].

7. **Administration**

The system should allow to define users and user groups in order to give them different roles and role-based access privileges. There should exist a register of all of the biobanks SOPs addressing all processes and activities of a biobank that should be followed by all of the personnel. The SOPs should be visible to all personnel but editable only by users with sufficient authorization. There should exist a register of all personnel that are that are performing any activities on the biological material of the biobank. There should

exist a register of ontologies that are used in the biobank to enforce the data standardization. The list should again be visible to all personnel but editable only by users with sufficient authorization. A list of all compatible printing software should be available, as it tells which barcode printers are supported by the system. The system should store information about the biobank. It should store the information about the ownership of the samples and data. The information about the biobank, which consists also of information about the owner and the biobank's address should conform to the MIABIS form for Data describing Biobank as shown in section 2.2.5. It should be publicly available. A document with the information relevant to biobank activities, processes and procedures should be made public as well. Lastly, the system should store documents that demonstrate the conformity with the GDPR and the ISO 20387:2018.
The system should:

7.1. Allow to define users, user groups and roles.

7.2. Have a register of all predefined SOPs (all personnel can see it; limited editing access).

7.3. Have a register with the identities of all personnel that are performing any activities on the biological material.

7.4. Have a register of controlled vocabularies (ontologies) to enforce data standardization and control (all personnel can see it; limited editing access).

7.5. Have a list of the compatible printing software (for barcodes).

7.6. Store information about ownership of the samples and data.

7.7. Publicly store information about the biobank using the MIABIS form Data describing Biobank.

7.8. Store a document with the information relevant to biobank activities, processes and procedures.

7.9. Store a document that demonstrates the conformity with the GDPR.

7.10. Store a document with the range of activities for which the biobanks conforms with the ISO 20387:2018.

— 4 —
Analysis

In this chapter, we first present some open source solutions that offer management of biobank information in section 4.1. We then give the comparison of the systems by analyzing the results from section 4.2 in section 4.3. To make it more coherent we analyze the requirements module by module as they were presented in chapter 3. Lastly, we comment on which open source solution would be the most appropriate to be used for running of a biobank with regard to our list of minimal requirements for a BIMS in section 4.4. We discuss what the chosen solution is lacking and how we could extend it.

4.1 Current open source solutions

In this section, we introduce some of the currently available open source solutions for the management of biobank information, that we then review in regard with the list of minimal requirements for a BIMS in section 4.2. We do not review all that are available. This is mainly for two reasons. Firstly, we do not review the applications that have a different focus from what we are aiming for (e.g. that the samples are the central points of the application). We have for example omitted CAISIS [8] since their application is organized around patients rather than around samples or individual studies. Even though it looked promising at first we have also omitted OBiBa [33] as it is focused on data management and has no coverage of the sample management (except from simple sample collection) and of storage, which are both inherent modules of a BIMS. Secondly, we could not review the applications that had neither documentation, a demo or a download link available.

4.1.1 ATiM

Advanced Tissue Management Application (ATiM) [1] is a software developed by the Canadian Tumour Repository Network (CTRNet) in collaboration with Canadian biobanks. As the CRTNet fosters studies into the determinants of cancer, the software has been designed primarily for the operation of tissue banks, where data is often focused around a patient instead of around samples as in BIMS. However, in the later steps of development, ATiM has been extended to be able to be used as a general BIMS. The information about ATiM can be found either on the online demo site [1] or on the documentation page [2].

A good feature of ATiM that we did not notice in any of the other systems is that the developers of the software made available a set of pre-made templates for SOPs and other

documents that are used in the biobank [44]. These could prove very useful to someone who is in the beginning steps of opening their own biobank.

4.1.2 Baobab LIMS

Baobab LIMS has been developed by University of the Western Cape (UWC) [43] and the South African National Biodiversity Institute (SANBI) [41] after recognizing the need for establishing and harmonizing national and regional biobank governance frameworks to address access to human and other ecological samples of academic interest in Africa. More information about the system can be found on the documentation page [3], on the wiki page [5], or in the latest documentation document [4].

4.1.3 OpenSpecimen

OpenSpecimen [40] is a web based open source informatics platform that is made primarily for managing of a biobank. It is developed in India by a company called Krishagni. It has an extensive documentation page [38] where we were able to find answers to almost all of our queries. The page seems to be in constant development as many of the topics we read had a very recent date of upload. Information about OpenSpecimen can also be found on an online forum with active users [37]. Aside from the open source solution, the company offers a payable enterprise edition with some additional features.

4.2 Results

The tables 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9 and 4.10 show the results of our reviewing of three open source solutions that we think could be appropriate to be used to manage a biobank. Each application was checked for each of the requirements. If the requirement is covered by the application it is marked by 'Yes', if it is not it is marked by 'No'. We use 'Partially' when only a part of the requirement is covered and 'N\A' when the information about this particular requirement was not found. The results from these tables are discussed in section 4.3.

While checking for required functionalities within each of the systems we tried to be as precise as possible. We believe we have found all of the existing functionalities. However, since the inspection of the systems was done by a single person we allow for the option that we might have missed some points and have thus graded some system unfairly.

4.3 Comparison of the systems

In this section we give the comparison of the ATiM, Baobab and OpenSpecimen module by module from the BIMS requirements' perspective.

1. **Sample management**

   1.1. **Sample registration**

   Out of all three applications, only Baobab enforces a globally unique ID for each sample entry, though it can be enabled in OpenSpecimen. We find this interesting as it is an important feature from the organisational point of view for samples in

	Requirement for Sample registration	*ATiM*	*Baobab*	*OpenSpecimen*
1.1.1	Enforce globally unique IDs for each sample entry	No	Yes	No
1.1.2	Enforce globally unique IDs for each donor	Yes	Yes	Yes
1.1.3	Generate and print barcodes, RFID and/or other machine readable labels	No	Partially	Yes
1.1.4	Be able to encode the personal data of human donors	No	Partially	Partially
1.1.5	Sample registration form should use documentation requirements from ISO 20387:2018	No	No	No
1.1.6	Extend the sample registration form with customizable fields	No	Partially	Yes
1.1.7	Have an option to mark whether the sample has been collected or acquired	No	No	No
1.1.8	Allow to choose a SOP that was used for collection/acquisition	Yes	Yes	Yes
1.1.9	Allow to flag and describe the deviation of the SOP for collection or acquisition process.	Yes	Partially	Yes
1.1.10	Import samples by scanning their existing barcodes	Yes	Yes	Yes
1.1.11	Bulk import samples	Yes	Yes	Yes
1.1.12	Allow linkage of aliquots to the parent sample	Yes	Yes	Yes
1.1.13	Allow linkage of derived samples to the parent sample	Yes	Yes	Yes
1.1.14	Allow to add notes on possible problems of the sample (bad quality for research, physical condition, leaks)	Yes	Partially	Yes
1.1.15	Add links to patient files	Yes	Yes	Yes
1.1.16	Link the sample to a study	Yes	Yes	Yes
1.1.17	Link to or upload a signed informed consent form (for human subjects)	Yes	Yes	Yes
1.1.18	Add information about the status of informed consent of a sample (signed, not-signed)	Yes	No	Yes
1.1.19	Validate the forms before accepting them	No	No	N/A

Table 4.1: Comparison of the chosen open source systems on the Sample registration module.

	Requirement for Sample tracking	*ATiM*	*Baobab*	*OpenSpecimen*
1.2.1	Have a secure, computer-generated, time-stamped audit trail of all changes/activities associated to a sample (from receiving a sample until distribution/destruction). It must include the date, time and who accessed the sample	Partially	Yes	Yes
1.2.2	Allow users to add additional information/notes of events (e.g. temperature logs) to specific samples	Yes	Partially	N/A
1.2.3	Allow users to change information about the status of informed consent of a sample (e.g. full or partial withdrawal of consent by the donor)	Yes	No	Yes
1.2.4	Have the procedure to delete all data from a certain donor	No	No	Partially
1.2.5	Anonymise the records of checked-out samples and store them for a predefined amount of time	Yes	No	Partially

Table 4.2: Comparison of the chosen open source systems on the Sample tracking module.

the storage. Differently, all applications enforce globally unique IDs for donors. As ATiM does not enforce globally unique IDs for samples it is of no surprise that it does not have support for generation and printing of machine readable labels. Baobab does have simple barcode generator, but users need to use their own software to create unique barcodes. OpenSpecimen supports this feature. Interestingly only one of the systems enables the encoding of personal data of human donors in a way that we think is appropriate for BIMS, which is encoding by pseudonymisation. This system is Baobab, where barcodes are used for IDs of donors. The users are not able to identify the donor unless they have access to the link where the documents (consent forms, metadata records) are stored. OpenSpecimen supports only encoding by anonymisation which is often not desired in medical research as it prevents from being able to contact the donor to give him or her results of the study. ATiM on the other hand offers no encoding of personal data at all. Here, the personal data are only hidden from users with no permission.

The ISO 20387:2018 is a rather new document published only recently (August 2018). Additionally, it is a standard that needs to be purchased, so it is according to our expectations that none of the systems support all of its particular requirements yet. Some requirements from this standard are quite common sense and have been used by biobanks already before its publication (e.g. the use of predefined SOPs and signing of transfer agreements prior to shipping of samples). All systems have a register of SOPs. Both ATiM and Baobab require a transfer agreement. As expected, none support the usage of documentation requirements from this standard that are explained in section 2.2.5. However, as OpenSpecimen is highly configurable the forms can be customized by adding the fields required by the ISO 20387:2018 and

	Requirement for Sample distribution and shipping	*ATiM*	*Baobab*	*OpenSpecimen*
1.3.1	Allow the user to create a list of samples for distribution	Yes	Yes	Yes
1.3.2	Allow the user to generate a list by uploading a list of samples	No	No	Yes
1.3.3	Have the functionality to process a sample request	Yes	Yes	Yes
1.3.4	Follow the SOP for distribution and shipping	Yes	Yes	Yes
1.3.5	Have the distribution form from the documentation requirements of ISO 20387:2018	No	No	No
1.3.6	Have the functionality to generate or store sample transfer agreements	Yes	Yes	No
1.3.7	Have the transfer form from the documentation requirements of ISO 20387:2018	No	No	No
1.3.8	Maintain the chain-of-custody	No	Partially	No
1.3.9	Enable to note the elements that may compromise the quality of the sample while in transfer	No	Yes	No
1.3.10	Note the proof of delivery when sample gets distributed	No	Yes	Yes
1.3.11	Store records of distributed samples for a predefined amount of time	Yes	Yes	Yes

Table 4.3: Comparison of the chosen open source systems on the Sample distribution and shipping module.

	Requirement for Sample destruction	*ATiM*	*Baobab*	*OpenSpecimen*
1.4.1	Allow complete removal of a subject and all of his samples and other information from the database	No	Yes	Partially
1.4.2	Have a SOP for destruction of a sample and associated data	No	Yes	No
1.4.3	Note the proof of destruction when the sample gets destructed	No	Yes	No
1.4.4	Store records of destructed samples including the reason for destruction for a predefined amount of time	No	No	No

Table 4.4: Comparison of the chosen open source systems on the Sample destruction module.

	Requirement for Study management	*ATiM*	*Baobab*	*OpenSpecimen*
2.1	Add a new study using the MIABIS form for Data describing Study	No	Yes	Yes
2.2	Add users to the study	N/A	Yes	Yes
2.3	Store the SOPs of the study if they differ from the biobank's	No	Partially	Yes
2.4	Store ethical approvals of the study	Yes	Yes	Yes
2.5	Store informed consents of the study	Yes	Yes	Yes
2.6	Show samples used in the study	Yes	Yes	Yes

Table 4.5: Comparison of the chosen open source systems on the Study management module.

	Requirement for Storage management	*ATiM*	*Baobab*	*OpenSpecimen*
3.1	Add and configure different kind of storage containers (freezers, trays, plates, boxes, samples)	Yes	Yes	Yes
3.2	Add and configure a hierarchy of storage containers	Yes	Yes	Yes
3.3	Add and configure different kind of boxes/samples in a freezer	Yes	Yes	Yes
3.4	Have a SOP for storage of different types of samples	Yes	Yes	Yes
3.5	The form should follow the form from the documentation requirements of ISO 20387:2018	Partially	No	No
3.6	Reorganize/move samples in existing freezers	Yes	Yes	Yes
3.7	Monitor storage conditions (e.g. deviations of temperature, humidity)	No	No	Yes

Table 4.6: Comparison of the chosen open source systems on the Storage management module.

	Requirement for System querying, reports and cataloging	*ATiM*	*Baobab*	*OpenSpecimen*
4.1	Allow end-users to search/query the biobank by using predefined queries	No	Yes	Yes
4.2	Allow end-users to search/query the biobank by using custom queries	Yes	Partially	Yes
4.3	Provide ability to save queries for future reuse	Yes	No	Yes
4.4	Allow end-users to find stored samples by barcodes or by IDs	N/A	Yes	Yes
4.5	Allow external researchers to search a sorted catalog of what is stored in the system	No	Yes	Yes
4.6	Allow end-users to generate predefined descriptive reports and summaries of what is stored in the system	Yes	Partially	Yes
4.7	Allow end-users to generate customized descriptive reports and summaries of what is stored in the system and to save them	No	Partially	Yes
4.8	Generate periodic reports of the system	No	Partially	Yes

Table 4.7: Comparison of the chosen open source systems on the System querying, reports and cataloging module.

	Requirement for Data import and export	*ATiM*	*Baobab*	*OpenSpecimen*
5.1	Allow end-users to import legacy data	Partially	Yes	Yes
5.2	Allow end-users to export data	Partially	Partially	Yes

Table 4.8: Comparison of the chosen open source systems on the Data import and export module.

	Requirement for Security	*ATiM*	*Baobab*	*OpenSpecimen*
6.1	Keep the personal information of donors separately from the coding key and ensure the secure storage, usage and follow up of the information	No	Yes	N/A
6.2	Support the use of secure, computer-generated, time-stamped audit trails. The computer should record what action was made, who made it and its date and time.	Partially	Yes	Yes
6.3	Should have regular backing-up of all data and the ability to restore the availability and access to personal data in the event of a physical or technical incident	Partially	Yes	Yes
6.4	Should be audited	No	Yes	Yes
6.5	Should allow only authenticated access	Yes	Yes	Yes
6.6	Should support secure connection protocols (e.g. SSL certificates)	Yes	Yes	Yes

Table 4.9: Comparison of the chosen open source systems on the Security module.

thus make it compliant. The sample registration form can be extended in Baobab as well, though not by an user, but rather by a software developer. Distinguishing between biosamples collected by the biobank and acquired from elsewhere is a new requirement from ISO so as could be expected none of the systems support it yet. ATiM and OpenSpecimen support both choosing a SOP that was used for the collection or acquisition process and flagging the deviation of the chosen SOP. Baobab supports referring to a SOP as well, while the module for flagging deviations is in development. The useful features of bulk importing samples and of importing samples by scanning their existing barcodes are supported by all of the systems. We were pleased to find out that all of the systems allowed to add notes on the possible problems of the sample. It is important to know if the sample's quality has been compromised as it might make it unusable for some or all research. As can be expected, all of the systems support basic features such as linking the aliquots and derivatives to the parent sample, linking the sample to the patient files (which are useful as additional data) and to a study. Since the informed consent is an important way to safeguard the rights of a donor we were a bit surprised to find that not all systems support all functions of an informed consent. It is not possible to add information about the status of informed consent in Baobab. None of the systems support the validation of the forms.

It is not easy to choose a winner of this module between Baobab and OpenSpecimen. Each has some advantages over the other. Baobab enforces globally unique IDs for each sample which is an essential organisational and tracking feature. It has a preferred option of encoding the personal data of human donors than OpenSpeci-

	Requirement for Administration	*ATiM*	*Baobab*	*OpenSpecimen*
7.1	Allow to define users, user groups and roles	Yes	Yes	Yes
7.2	Have a register of all predefined SOPs	Yes	Yes	Yes
7.3	Have a register with the identities of all personnel that are performing any activities on the biological material	No	Partially	Yes
7.4	Have a register of controlled vocabularies (ontologies) to enforce data standardization and control	Yes	Yes	Yes
7.5	Have a list of the compatible printing software	No	Partially	Yes
7.6	Store information about ownership of the samples and data	Yes	Yes	No
7.7	Publicly store information about the biobank using the MIABIS form Data describing Biobank	No	No	Yes
7.8	Store a document with the information relevant to biobank activities, processes and procedures	No	No	No
7.9	Store a document that demonstrates the conformity with the GDPR	No	No	No
7.10	Store a document with the range of activities for which the biobanks conforms with the ISO 20387:2018	No	No	No

Table 4.10: Comparison of the chosen open source systems on the Administration module.

men. On the other hand, OpenSpecimen has more customizable forms and enables editing the status of informed consent of the donor which is crucial for the donor rights as stated by the GDPR. As such, neither of the two systems can be chosen as the better one simply after overview of this module.

1.2. **Sample tracking**

The coverage of functionalities is not as good in the Sample tracking module. None of them is fully covered by all three of the systems. The support of audit trail is full in Baobab and OpenSpecimen but only partial in ATiM. As already noted in the previous module, Baobab does not support changing the status of informed consent of the sample. This is a big drawback of the system as it is imperative for the compliance with the GDPR. The functionality is covered by both other systems. Interestingly though, it is not possible to delete all data from a certain donor in either of the systems. This is not in accordance with an important right of a data subject by the GDPR, the right to be forgotten. According to it, the donor should have the right to ask for all of his or her personal information, including the samples to be removed from the biobank. Due to the nature of research in biobanks, they

can often get exempted from this obligation so they may keep samples and data that are needed for research. OpenSpecimen allows to anonymise all personal data of a donor. The associated samples thus become de-identifiable which is acceptable for the GDPR. In Baobab, an IT administrator needs to manually delete the data in the backend but the history of the event is recorded, so this is not sufficient. ATiM does not offer this functionality at all and we think that it is imperative that it gets added. Only ATiM and OpenSpecimen can anonymise the records of checked out samples.

The biggest drawback of OpenSpecimen against other two systems in this section is that we could not find information about being able to add notes of events on specific samples in storage. But as this functionality is of a much lesser importance than the full audit trail of the samples that is missing in ATiM and compliance with the right to be forgotten of GDPR which is lacking in Baobab, OpenSpecimen is a clear winner of this module.

1.3. **Sample distribution and shipping**

All systems enable basic functionalities of distribution which are creating a list of samples for distribution and processing a sample request. Additionally, OpenSpecimen supports generating the list of samples by uploading the pre-made list while ATiM and Baobab support sample transfer agreements. These are rather important as they include the rules on how the recipient of the samples and the associated data will treat and protect the personal information of donors.

Neither of the ATiM and OpenSpecimen is very good for the transfer of the sample. ATiM actually has no support for transfer at all. Once the samples have been selected for distribution all that is possible is to manually change their status to 'shipped'. This will at least keep all of their information in the system as the last feature of this module requires. The drawback of OpenSpecimen is its prerequisite for shipping. It is that the shipping and the receiving site are using the same instance of OpenSpecimen. This allows for interoperability of only a small number of biobanks which is not desired. If the prerequisite is met, OpenSpecimen supports noting the proof of delivery of samples. Only Baobab supports the chain-of-custody during the transfer of the sample and even that only partially as dedicated couriers are used. It is an important feature for maintaining the sample quality so we are surprised that none of the systems fully supports it. Moreover, it is an requirement from ISO.

The winner of this module is Baobab as it offers a full sample shipment module and covers the widest array of transfer features.

1.4. **Sample destruction**

Neither of ATiM or OpenSpecimen has support for this module. As discussed before, OpenSpecimen enables only anonymisation of personal information meaning that all other information and samples still stay in the system. Baobab is an evident winner of this module as it supports almost all of its features. The only feature it does not support is the storage of records of destructed samples. Once a record is deleted it is removed entirely. We are however not sure if the removal of a subject and his or her samples and information can be initiated and done from the application itself or if the IT administrator needs to do it manually in the backend.

2. **Study management**

We found MIABIS guidelines a very convenient mode of facilitating harmonization of biobanks. As it is freely available we hoped that at least some of the open source systems will support it. We were pleased to find that Baobab collects all information that is prescribed in the Data describing Study in section 2.2.5. We were also satisfied to find it in the documentation of OpenSpecimen. Even though it is not a part of basic installation, the documentation already includes a guide on how to change the generic data collection forms that come with the installation of OpenSpecimen so that they will comply with the guidelines of MIABIS. Most fields only need to be renamed while a few need to be added. ATiM does not include MIABIS. All of the applications support basic features of this module which are storing ethical approvals and informed consents of the study and the ability to show the list of samples used in the study. In ATiM it is not possible for the study to upload SOPs that differ from the biobanks'. It is of course encouraged for all of the studies and biobanks to use the same SOPs as it makes interoperability of biobanks more trivial. But this is realistically hard to expect, so allowing the study to add their own SOPs is practical. In Baobab SOPs can be stored in the system and while they cannot be linked directly from the study, notes of their usage can be made in the *description* field of the study. We were more surprised that we could not add the users to the study in ATiM. Users are added to the study to enable them to see the data of the study and to limit their access to other information within the biobank. It is an important security feature of a BIMS.

Baobab and OpenSpecimen both support all features of this module. However, as Baobab has an already implemented MIABIS form for Data describing Study and a sufficient support for SOPs that differ from a biobank's, we would mark it as the winner of this module but only by a small margin.

3. **Storage management**

   All systems have a good coverage of features of this module. They all support adding different kinds of storage containers and defining their hierarchy, configuring boxes and samples within a certain freezer and moving the samples around storage. Additionally, the documentation of samples within storage in ATiM covers some of the documentation requirements from ISO 20387:2018 while OpenSpecimen supports integration with automated freezers to monitor storage conditions (e.g. changes in temperature in humidity). Monitoring of storage conditions has a higher value than a partial compliance to storage documentation requirements of ISO as it has a higher impact on maintaining the sample quality and integrity. Especially as it is possible to make OpenSpecimen compliant to ISO because of its high configurability.

   As such, OpenSpecimen is a winner of this module with ATiM and Baobab very close behind.

4. **System querying, reports and cataloging**

   OpenSpecimen has a very broad support of querying, reporting and cataloging as it covers all of the features of this module. The coverage is not as good in ATiM and Baobab. The users of Baobab can query the system by using predefined queries while they are limited when using custom queries and thus cannot save them. In ATiM, users cannot query the system by using predefined queries but only by custom ones. Even though they can save their queries for future use this functionality is rather crude as the users cannot save an

actual query but can instead only write down and save the steps of a query. Samples can be found by barcodes (which serve as IDs) in Baobab. We could not find information about this feature for ATiM. Surprisingly, ATiM does not provide a catalog of what is stored in the system. This is a rather big disadvantage as external researchers cannot look for samples that they could use in their research. Catalog exists in Baobab. Moreover, the reporting is rather limited in both systems. In ATiM, it only supports generation of predefined reports and not of customized or periodic reports, while the reporting module in Baobab is still in development.

OpenSpecimen is an obvious winner of this module. Once the development on the reporting module in Baobab is finished, it will become comparable to OpenSpecimen.

5. **Data import and export**

   OpenSpecimen has support of both import and export. As ATiM can export and import only dumps for and from other ATiM systems and the exporting is still under development in Baobab, OpenSpecimen is a clear winner of this module.

6. **Security**

   Only Baobab keeps personal information of donors separately from the coding key, which is an important security measure. Baobab and OpenSpecimen have a full support of computer-generated audit trails while ATiM audits only changes made to user data. This is a drawback of ATiM as audit trails are an essential security measure. Both Baobab and OpenSpecimen support regular backups, while in ATiM only manual backups are possible. As for the audited systems, OpenSpecimen gives some pointers on how to achieve the computer system validation by National Cancer Institute [39]. Baobab allows the capture of accreditation details of the laboratory. All systems support secure connection protocols and authenticated access.

   Since Baobab supports all of the functionalities of this module, especially important ones such as keeping the coding key in a safe place, audit trails and regular backups, it is the winner of this module.

7. **Administration**

   Allowing to define users, user groups and roles and a register of controlled vocabularies are the only requirements aside from having a register of SOPs that all three applications support. OpenSpecimen has an useful list of compatible barcode printing software, while ATiM claims that the biobank is the owner of all samples within it. In Baobab, the client who proposes a study is the owner of the samples and data that belong to the study. Requirements for the register of all personnel that are performing activities on biological material, for publicly storing information about the biobank, for storing a document with information relevant to biobank activities, processes and procedures and for storing a document with the range of activities for which the biobank conforms with ISO 20378:2018 are the requirements of this standard. As discussed before this is a payable and relatively new standard and as such it is of no surprise that these requirements are weakly or not supported by any of the applications yet. None of the applications store the document that demonstrates the conformity with GDPR either. This would be an useful feature to add in order to avoid hefty GDPR fines.

All systems are very comparable in this module as they all offer the support of the same base features of this module. It is only because OpenSpecimen has a list of printing software that it makes it a winner of this module.

4.4 Discussion

The modular comparison of the systems did not produce one system that would be evidently better than the other two. Comparing how many times each system comes out as a winner of a module, we realized that Baobab and OpenSpecimen are equal and both better than ATiM (which is never a winner of a module). This number does however not tell us everything. It does not mean that we could choose either Baobab or OpenSpecimen as an equally sufficient BIMS. We also need to look at importance and quality of the features that each of the systems supports or lacks. This is how we chose OpenSpecimen as a more appropriate BIMS to manage a biobank.

A considerable reason why we make such a decision is that Baobab does not support the full functionalities of an informed consent. It is not surprising as the software is developed in Africa, but as we are located in the European Union, the compliance to GDPR is of utmost importance when handling sensitive data such as personal information. Being able to change the status of an informed consent, to mark it as withdrawn and thus make the donor's samples unavailable facilitates following GDPR and reduces the risk of using samples of a donor that has forbidden it. Furthermore, OpenSpecimen enables integration of automated freezers while Baobab does not. This facilitates the control over the conditions in which samples are stored and helps controlling the continuous sample quality over a long period of storage. Another thing that Baobab lacks is the ability to encode personal data by anonymisation. It has no option to make data de-identifiable. Lastly, OpenSpecimen has a better quality of certain features. It supports barcode generation. It has a more comprehensive audit-trail, while the team at Baobab is currently still improving this feature. Baobab's reporting, importing and exporting modules are in development while all are fully functional in OpenSpecimen.

There are however some features that Baobab has but OpenSpecimen lacks. It enforces globally unique IDs for each sample entry, though this can be enabled in OpenSpecimen as well. We were not able to find information about whether OpenSpecimen keeps personal information of donors separately from the coding key while Baobab does. Bigger disadvantages of OpenSpecimen over Baobab are that it does not have shipping and destruction modules. We are however not sure if the removal of a donor and all of his or her samples and information from the database can be made from the software or if the IT administrator has to do it. If it is the second option, it is not a disadvantage anymore as the same can be done in OpenSpecimen. The shipping module is a big downside of OpenSpecimen although not as big as not being able to control informed consents in Baobab as this is an essential feature because of the GDPR. The shipping module of Baobab is very advanced as it goes so far as to even organize kits, into which the samples are packed during transport.

Yet another reason why we prefer OpenSpecimen over Baobab is the documentation of the systems. OpenSpecimen has an interactive documentation page on which new topics are added almost weekly. It has a functioning search function and a forum with active users where the questions can be asked. On the other hand, the currently available documentation from Baobab is rather insufficient. Quite some items need more comprehensive explanations. By comparison with the documentation of OpenSpecimen, it was much more difficult to find information on

the documentation page of Baobab. We even had to contact the team to help us with checking of existence of requirements from our list as we were not able to do it only by the help of the documentation. The responses however were fast and informative. We were told that more recent functionality additions have not been documented yet. When the functionalities that are in development get finalized and when the documentation gets updated it will be difficult to say whether OpenSpecimen is still more appropriate for the management of a biobank.

— 5 —
Conclusion

and In this work, we aimed to find an open source BIMS solution that is appropriate to manage a biobank. Many smaller research groups with lesser fundings cannot afford to use one of the commercial solutions as these tend to be very expensive. They may also not have funding, time or expertise to develop their own systems to manage a biobank. Therefore solutions should exist that such research groups could use in order to facilitate research. However, these solutions must be comparable to the commercial solutions in what features they offer. The biobanks that use them should be competitive with other biobanks in that they can ensure the proper quality of the biosamples that they collect. An adequate solution needs to comply at least with the requirements of the regulations and standards such as GDPR and ISO 20387:2018 in order to avoid the possibility of fines. Preferably, it should also support many other guidelines and functionalities.

In order to be able to compare different open source BIMS solutions we have created a list of minimal requirements that any adequate BIMS should support. Using this list we could compare the coverage of the requirements of each open source system and choose the best option. We have merged many important documents into the list of minimal requirements for a BIMS. We have chosen a selection of essential or commonly used regulations, standards and guidelines and have combined them with the knowledge of general functionalities that we have obtained from studying biobank-related papers and articles. General Data Protection Regulation (GDPR) is regulation on the protection of natural persons in relation to the processing of personal data. Since biobanks process mainly just personal data it is an essential regulation to take into account. General requirements for biobanking ISO 20387:2018 gives general knowledge about how all processes and procedures in a biobank should be run. The parts on confidentiality, structural requirements and process requirements are particularly informative. This standard emphasizes the use of SOPs for all steps in the life cycle of a biological material and for all other processes within a biobank. It highlights the need of harmonization of data in order to enable easier interoperability between different biobanks. Harmonization is also a focal point of MIABIS guidelines. FDA 21 CFR Part 11 is a regulation on electronic records and electronic signatures that can be used in lieu of their paper counterparts. The requirements that come particularly into account for building a BIMS are: the validation of the system, protection of records, role-privileged limited access and audit trail. ISBER Best Practices is an information system evaluation checklist of effective practices for the management of collections of biological material. We included all of the best practices that are relevant for our list of minimal requirements for a BIMS.

Three open source solutions, namely ATiM, Baobab and OpenSpecimen were compared

using the list. OpenSpecimen was chosen as the most adequate solution to be used as a BIMS for a biobank. This system is however not perfect and lacks support for a number of features. Most disappointingly, it lacks almost all functionalities of a shipping module. Even with current support, two biobanks need to use the same instance of OpenSpecimen to be able to use the existing functionalities. This is very limiting in the aspect of promoting interoperability of biobanks. There is almost no support for the removal of a donor and all of his or her samples and other information. This is very important for the right to be forgotten of GDPR. Currently, only anonymisation is supported, but the data should get removed from the system and not just de-identified. Support for the destruction of a sample, such as a proof of destruction is also missing. The encoding of personal information of a donor should enable pseudonymisation. This allows to protect the sensitive information of a donor while keeping the connection to be able to provide him or her with the results of the research. If samples cannot be linked to a person, this person also looses a right to be forgotten or to withdraw a consent. The coding key used for pseudonymisation should be stored separately from the personal information and have a secure storage. OpenSpecimen also has no support for the documentation requirements of ISO 20387:2018. Moreover, it lacks both a document that demonstrates conformity with GDPR and a document with the range of activities for which the biobanks conforms with the ISO 20387:2018. Both are useful to have in case of a control of the system.

However, OpenSpecimen is available under an open source license that allows users to modify the source code. As such, plugins for all of the lacking functionalities could be developed as an interesting future work. This would make the system fully usable to manage a biobank. Additional future work could be to look into even more guidelines. One could see if there are any features mentioned in these guidelines that could be added to our list. This way, further credibility of the list could be reached. Good places to start would be the Recommendations for Repositories [24] by ISBER and Guidelines for Human Biobanks and Genetic Research Databases [36] and Guidelines for Biological Resource Centres [34] by OECD, an Organization for Economic Co-operation and Development [35].

Even though there are things that still need to be done we have made a big step in the right direction. We have made a comprehensive overview of many important standards, regulations and guidelines and combined them into a list of minimal requirements. This list is a strong contribution to the world of biobanking as it can be used to check the appropriateness of any commercial or open source BIMS solution. We have made an evaluation of three interesting open source solutions and highlighted one to ease the decision of choosing a open source BIMS. As a future work, this solution could be extended with missing features and applied to ease the management of a new biobank.

— A —
Appendix

Attribute code	*Attribute name*	*Allowed values*	*Description*
MIABIS-2.0-01	ID	Text	Textual string of letters starting with the country code followed by the underscore '_' and post-fixed by a biobank ID or name specified by its juristic person
MIABIS-2.0-02	Acronym	Text	Textual string of short name in use for the biobank. If applicable
MIABIS-2.0-03	Name	Text	Textual string of letters denoting the name of the biobank in English
MIABIS-2.0-04	URL	Text	Textual string of letters with the complete http-address for the biobank
MIABIS-2.0-05	Juristic Person	Text	Textual string of letters denoting the juristic person e.g. a university, concern, county council etc. for the biobank
MIABIS-2.0-06	Country	Text	ISO-standard (3166 alpha2), two letter code for the country of the biobank
MIABIS-2.0-07	Contact Information	Structured data	Contact information for the contact person of the biobank as defined in MIABIS-2.0-07
MIABIS-2.0-08	Description	Text	Textual string of letters with a description about the biobank in English

Table A.1: Data describing biobank

Attribute code	*Attribute name*	*Allowed values*	*Description*
MIABIS-2.0-01	ID	Text	Sample Collection ID that also links the sample collection to the hosting biobank or study
MIABIS-2.0-02	Acronym	Text	Short name in use for the sample collection
MIABIS-2.0-03	Name	Text	The name of the sample collection in English
MIABIS-2.0-08	Description	Text	A description of the sample collection. Recommendation max. 2000 char.
MIABIS-2.0-09	Sex	Text list	The sex of the individuals in the sample collection. Can be one or more of the following values: Male, Female, Unknown, Undiferrentiated
MIABIS-2.0-10	Age Low	Integer	Age of youngest sample donor at time of sample donation
MIABIS-2.0-11	Age High	Integer	Age of oldest sample donor at time of sample donation
MIABIS-2.0-12	Age Unit	Text list	Unit defining Age Low and Age High. Can be one of the following values: years, months, weeks, days
MIABIS-2.0-13	Data categories	Text list	The data catagories from which data is available. Can be several values described in MIABIS-2.0-13
MIABIS-2.0-14	Material type	Text list	The biospecimen saved from a biological entity for propagation e.g. testing, diagnostics, treatment or research purposes. Can be several values as defined in MIABIS-2.0-14
MIABIS-2.0-15	Storage temperature	Text list	The long term storage temperature at which the samples are stored after preparation. Can be several values as defined in MIABIS-2.0-15
MIABIS-2.0-16	Collection type	Text list	The type of the sample collection. Can be several values as defined in MIABIS-2.0-16
MIABIS-2.0-17	Disease	Structured data	The disease of main interest in the sample collection, if any. Can be several values as defined in MIABIS-2.0-17
MIABIS-2.0-07	Contact information	Structured data	Contact information for the contact person of the sample collection as defined in MIABIS-2.0-07 and MIABIS-2.0-23

Table A.2: Data describing sample collection

Attribute code	*Attribute name*	*Allowed values*	*Description*
MIABIS-2.0-01	ID	Text	The unique ID or acronym of the study
MIABIS-2.0-02	Name	Text	The name of the study in English
MIABIS-2.0-08	Description	Text	A description of the study aim. Recommendation max. 2000 char.
MIABIS-2.0-18	Principal Investigator	Text	The name of the person responsible for the study e.g. the principal investigator
MIABIS-2.0-07	Contact information	Structured data	Contact information for the contact person of the study as described in MIABIS-2.0-07 and in MIABIS-2.0-23
MIABIS-2.0-19	Study design	Text list	The design of the study. Can be one or several values as described in MIABIS-2.0-19
MIABIS-2.0-09	Sex	Structured data	The sex of the study participants. Can be one or more of the following values: Male, Female, Unknown, Undifferentiated as described in MIABIS-2.0-09
MIABIS-2.0-10	Age Low	Integer	Age of the youngest study participant at time of inclusion
MIABIS-2.0-11	Age High	Integer	Age of the oldest study participant at time of inclusion
MIABIS-2.0-12	Age Unit	Text list	Unit defining Age Low and Age High. Can be one of the following values: years, months, weeks, days
MIABIS-2.0-13	Data categories	Text list	The type of data that is associated with the samples in the study. Can be several values as described in MIABIS-2.0-13
MIABIS-2.0-14	Material type	Text list	The biospecimen saved from a biological entity for propagation e.g. testing, diagnostics, treatment or research purposes. Can be several values as described in MIABIS-2.0-14
MIABIS-2.0-20	Total number of participants	Integer	Total number of individuals recruited to the study
MIABIS-2.0-21	Total number of sample donors	Integer	Total number of individuals with biological samples in the study
MIABIS-2.0-22	Inclusion criteria	Text list	Information on type of parameters that determine which individuals will become study participants. Can be several values as described in MIABIS-2.0-22

Table A.3: Data describing study

Requirement from ISO 20387	*Documentation examples*
timestamp	collection time and/or date in a standard format preferably according to ISO 8601
collection site	geographical data of collection site (e.g. coordinates)
	host/source description (e.g. farm, hospital, animal, human, forest, field) environmental data of collection site
provider	name, address, code
	consent information, authorization, permission
	historical data, provenance
biological material/organic entity identification or characterization	consent information
	anonymisation/pseudonymisation
	taxonomy
	phenotypic data
	clinical data, diagnosis, treatments
	biometric data
	omics data
	epidemiological data
	life style data: smoking status, diet etc.
	demographic data
	unique identifier
	sample/isolate history
collection method	method of sampling
	primary container type
	additives, stabilizers
	final concentration of the sample
	storage conditions prior to shipment
specific properties	infectiousness
	biosafety information, radioactivity/radiation
	transgenic/chimera/genetically modified etc.

Table A.4: Documentation requirement for acquisition of biological material

Requirement from ISO 20387	*Documentation examples*
long-/short-term storage conditions	type of storage, freezer or cold room
	temperature
	timestamp
	humidity
	exposure to radiation (e.g. light)
	duration
	access procedure
	safety
	container type
	traceability

Table A.5: Documentation requirement for storage of biological material

Requirement from ISO 20387	*Documentation examples*
compliance with applicable regulatory and ethical requirements	as applicable
verification of biological material and of associated data	as applicable
contractual information	authorized representatives of the contracting parties
	title of the project
	obligation of recognizing or citing the biobank
	requirement for feedback of information and scientific results

Table A.6: Documentation requirement for distribution and disposal of biological material

Requirement from ISO 20387	*Documentation examples*
mode of transportation/shipment specifications	UN number (e.g. UN 3373), packing instructions (e.g. PI650), prohibitions (e.g. radiation)
temperature during transport	min./max. specifications
temperature at reception	min./max. specifications
transport start and end time and date in a standard format according to ISO 8601	max. specifications
other requirements	humidity, light, maximum shipping time, climate/season

Table A.7: Documentation requirement for transport of biological material

Bibliography

[1] Advanced Tissue Management Application (ATiM). https://www.ctrnet.ca/resources/atim. Accessed: 2019-05-20.

[2] CTRNet Documentation Centre. http://www.ctrnet.ca/mediawiki/index.php?title=Main_Page. Accessed: 2019-05-20.

[3] Baobab LIMS Documentation page. https://baobab-lims.readthedocs.io/en/latest/. Accessed: 2019-05-30.

[4] Baobab LIMS Documentation, latest release 1.3. https://buildmedia.readthedocs.org/media/pdf/baobab-lims/latest/baobab-lims.pdf. Accessed: 2019-05-30.

[5] Baobab LIMS Wiki. https://github.com/BaobabLims/baobab.lims/wiki. Accessed: 2019-05-30.

[6] Biobank operational module. http://bibbox.bbmri-eric.eu/about/biobank-operational-module/. Accessed: 2019-03-08.

[7] Biobanking: FAQs. Technical report, European Cancer Patient Coalition. Accessed: 2019-02-23.

[8] Caisis - Open Source Cancer Data Management System. http://www.caisis.org/. Accessed: 2019-05-08.

[9] CFR - Code of Federal Regulations Title 21. https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=11.1. Accessed: 2019-05-08.

[10] Title 21 CFR Part 11. https://en.m.wikipedia.org/wiki/Title_21_CFR_Part_11. Accessed: 2019-05-08.

[11] Tutorial on 21 CFR Part 11 - Electronic Records and Electronic Signatures. http://www.labcompliance.com/tutorial/part11/default.aspx?sm=d_c. Accessed: 2019-05-08.

[12] Geoffrey M Cooper. *The Cell: A Molecular Approach. 2nd edition*. Sinauer Associates, 2000.

[13] BBMRI-ERIC Common Service ELSI. The EU General Data Protection Regulation, Answers to Frequently Asked Questions . 2017.

[14] L. Mook E.P.A van Iperen, D. van Enckevort. Minimal requirements document bims. 2016.

[15] Council of the European Union European Parliament. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *OJ L 119*, pages 1 – 88, 2016.

[16] Martin N Fransson, Emmanuelle Rial-Sebbag, Mathias Brochhausen, and Jan-Eric Litton. Toward a common language for biobanking. *European Journal of Human Genetics*, 23:22 – 28, 2015.

[17] Geoffrey S. Ginsburg and Jeanette J. McCarthy. Personalized medicine: revolutionizing drug discovery and patient care. *Trends in Biotechnology*, 19(12):491 – 496, 2001.

[18] Margaret A. Hamburg and Francis S. Collins. The path to personalized medicine. *New England Journal of Medicine*, 363(4):301–304, 2010.

[19] Jarkko Hyysalo, Anja Keskinarkaus, Gavin Harper, and Jaakko J. Sauvola. Architecture enabling service-oriented digital biobanks. In *HICSS*, 2017.

[20] International Classification of Diseases (ICD) . `https://www.who.int/classifications/icd/en/`. Accessed: 2019-05-02.

[21] About ISBER. `https://www.isber.org/page/About`. Accessed: 2019-05-08.

[22] ISBER Informatics Working Group. `https://www.isber.org/page/INF`. Accessed: 2019-05-08.

[23] ISBER Published documents. `https://www.isber.org/page/ISBERPD`. Accessed: 2019-05-08.

[24] ISBER Recommendations for Repositories, Fourth edition. `https://www.isber.org/page/BPR`. Accessed: 2019-06-01.

[25] ISO. Biotechnology – Biobanking – General requirements for biobanking. ISO 20387:2018(E), International Organization for Standardization, Geneva, Switzerland, August 2018.

[26] Tae Hoon et al. Jang. Cryopreservation and its clinical applications. *Integrative medicine research*, 6(1):12 – 18, 2017.

[27] AK Macleod, DCM Liewald, MM McGilchrist, AD Morris, SM Kerr, and DJ Porteous. Some principles and practices of genetic biobanking studies. *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology*, 33:419–25, 03 2009.

[28] Roxana Merino-Martinez, Loreana Norlin, David van Enckevort, Gabriele Anton, Simone Schuffenhauer, Kaisa Silander, Linda D. Mook, Petr Holub, Raffael Bild, Morris A. Swertz, and Jan-Eric Litton. Toward global biobank integration by implementation of the

minimum information about biobank data sharing (miabis 2.0 core). *Biopreservation and biobanking*, 14(4):298–306, 2016.

[29] Lucia Migliore and Fabio Coppede. Genetics, environmental factors and the emerging role of epigenetics in neurodegenerative diseases. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 667(1):82 – 97, 2009.

[30] F. John Mills and Andrew Brooks. Information Systems for Biobanks/Biorepositories: Handling Information Associated with Compliant Sample Management . *Biopreservation and biobanking*, 8(3), 2010.

[31] Heimo Müller, Nicolas Malservet, Philip Roy Quinlan, Robert Reihs, Matthieu Penicaud, Antoine Chami, Kurt Zatloukal, and George Dagher. From the evaluation of existing solutions to an all-inclusive package for biobanks. In *Health and technology*, 2017.

[32] Natural person. `https://en.m.wikipedia.org/wiki/Natural_person`. Accessed: 2019-02-23.

[33] OBiBa - Open Source Software for Epidemiology. `https://www.obiba.org/`. Accessed: 2019-05-08.

[34] OECD Best Practice Guidelines for Biological Resource Centres. `http://www.oecd.org/sti/emerging-tech/oecdbestpracticeguidelinesforbiologicalresourcecentres.htm`. Accessed: 2019-06-01.

[35] OECD Home. `http://www.oecd.org/`. Accessed: 2019-06-01.

[36] Guidelines for Human Biobanks and Genetic Research Databases (HBGRDs). `http://www.oecd.org/sti/emerging-tech/guidelines-for-human-biobanks-and-genetic-research-databases.htm`. Accessed: 2019-06-01.

[37] OpenSpecimen Forums. `http://forums.openspecimen.org/`. Accessed: 2019-05-20.

[38] OpenSpecimen support center. `https://openspecimen.atlassian.net/wiki/spaces/CAT/overview`. Accessed: 2019-05-20.

[39] Validation in OpenSpecimen. `https://openspecimen.atlassian.net/wiki/spaces/CAT/pages/1115985/Is+OpenSpecimen+CFR+Part+11+complaint`. Accessed: 2019-05-20.

[40] OpenSpecimen Platform. `https://www.openspecimen.org/`. Accessed: 2019-05-20.

[41] The South African National Biodiversity Institute (SANBI). `https://www.sanbi.org/`. Accessed: 2019-05-30.

[42] SNOMED. `http://www.snomed.org/`. Accessed: 2019-05-02.

[43] University of the Western Cape. `https://www.uwc.ac.za/Pages/default.aspx`. Accessed: 2019-05-30.

[44] Products and Services. https://biobanking.org/webs/other_products. Accessed: 2019-05-20.