# Sample Quality as Basic Prerequisite for Data Quality: A Quality Management System for Biobanks

Christiane Hartfeldt[1(✉)], Verena Huth[1], Sabrina Schmitt[2], Bettina Meinung[4], Peter Schirmacher[3], Michael Kiehntopf[4], Cornelia Specht[1], and Michael Hummel[1]

[1] German Biobank Node, Charité - Universitätsmedizin Berlin, Berlin, Germany
christiane.hartfeldt@charite.de
[2] BioMaterialBank Heidelberg, Heidelberg University Hospital, Heidelberg, Germany
[3] Institute of Pathology, Heidelberg University Hospital, Heidelberg, Germany
[4] Integrated Biobank Jena, Jena University Hospital, Jena, Germany

**Abstract.** Artificial intelligence will undoubtedly shape technological developments in the next years – affecting biobanks as well. First concepts already exist for the possible application of AI in biobank processes. Data quality, which depends on sample quality, plays the decisive role here. For this reason, high and also overarching quality standards are important to prepare biobanks for these technological innovations. In addition, the requirements for sample and data quality will be significantly affected by the EU regulation for in vitro diagnostics as its transition period ends in 2022. The demand for human biospecimens will consequently rise.

In order to meet such requirements, the German Biobank Node (GBN) and the German Biobank Alliance (GBA) have established a quality management programme for German biobanks which includes e.g. an extensive quality manual and so-called "friendly" (cross-biobank) audits. The following article describes these developments with regard to their relevance for future biobank workflows using AI methods.

**Keywords:** Biobank · Quality · Machine learning

## 1 Introduction and Motivation

Assume it is 2029. The first approval for targeted treatment decisions based on image analysis is in place. Recognition of a distinct histomorphological pattern determines the most effective drug. A new era in precision oncology has begun – requiring considerations regarding the quality of the tissue specimens including their pre-analytical condition. Luckily, grounds for quality assessment and assurance have been laid in 2018/19 facilitating the introduction of morphology based treatment decision systems without any delay. Thanks to the pioneer work of the biobanking community in BBMRI-ERIC [1] and its 21 national nodes, a quality


A. Holzinger et al. (Eds.): AI & ML for Digital Pathology, LNAI 12090, pp. 89–94, 2020.
https://doi.org/10.1007/978-3-030-50402-1_5

system has been established allowing harmonised handling of tissue samples to guarantee comparable and reliable results irrespective of local conditions. The following text describes the components and the establishment of this quality system in Germany in cooperation with BBMRI-ERIC.

## 2 Glossary

*Biobanking*: describes biobank processes such as the reception, processing, storage, transport, retrieval, and distribution of biospecimens and associated data for biomedical research.

*German Biobank Node and German Biobank Alliance*: Led by the German Biobank Node (GBN), the German Biobank Alliance (GBA) is a network consisting of 18 biobank sites and two IT centres in Germany [2]. The alliance partners establish quality standards for biospecimens and related data from different biobanks nationally and internationally available for biomedical research. GBN is the umbrella organisation for university biobanks in Germany. Since 2017 the biobanks of GBA and GBN are funded by the Federal Ministry of Education and Research (BMBF). GBN serves in addition as the German national node for BBMRI-ERIC.

*BBMRI-ERIC*: is a pan-European infrastructure of national biobank networks [1]. The abbreviation BBMRI-ERIC stands for "Biobanking and Biomolecular Resources Research Infrastructure – European Research Infrastructure Consortium". BBMRI-ERIC's aim is to facilitate access to high-quality biospecimens and associated data as research resources.

*Sample Locator*: is a web-based search tool enabling scientists to perform cross-biobank searches for biosamples and related data meeting relevant criteria. Currently, GBA biobanks are connected and the IT infrastructure is prepared to include further national or European biobanks. It was developed by GBN in cooperation with the German Cancer Research Center in Heidelberg (DKFZ).

## 3 State-of-the-Art

Artificial intelligence will shape technological developments in the next years – affecting biobanks as well. First concepts already exist for the possible application of AI in biobank processes. Data quality, which depends on sample quality, plays the decisive role here.

For this reason, high and also overarching quality standards are important to prepare biobanks for these technological innovations. In addition, the requirements for sample and data quality will be significantly affected by the EU regulation for in vitro diagnostics as its transition period ends in 2022. The demand for human biospecimens will consequently rise.

The German Biobank Node (GBN) [2] and the German Biobank Alliance (GBA) have established a quality management programme for German biobanks which includes e.g. an extensive quality manual and so-called friendly (cross-biobank) audits.

### 3.1 The Quality Management Concept of GBN

The main objective of the activities by GBN and GBA in the field of quality management is the implementation and maintenance of a uniform quality management system (QMS) at all GBA sites. This is accomplished by means of a QM manual of generic biobank documents, an uniform QM software for the management of documents and processes, a cross-GBA audit system, ring trials for the review of sample-processing steps and constant review of the satisfaction of biobank users with the offered services.

### 3.2 QM Manual

The GBN QM manual consists of generic biobank-specific standard operating procedures. It is based on several standards e.g. ISO 9001, ISO 15189, ISO 20387 and also contains documents for the management level. The manual has been published open access [3].

### 3.3 Software Solution

The central QM software solution for all GBA biobanks provides a digital interface for various quality applications and reduces the documentation effort by continuous improvement of processes.

### 3.4 Friendly Audits

Regular audits are an indispensable prerequisite to continuous review and improvement of biobank processes. All applied standards require audits to ensure compliance and improve the quality of biospecimens and related data. The QM core team of GBA developed an audit programme to perform cross-biobank audits based on applied standards and the audit specific standard ISO 20387. During a "friendly audit", auditors from other GBA biobanks audit each participating biobank once a year. As a result, auditors gain a new perspective and deeper understanding of the standards' requirements. For the biobanks, the audit programme supports preparation for external competence evaluation (certification, accredation) – which by 2020 all participating biobanks are likely to have passed. Other national networks such as the German Centers for Health Research (DZG) and the German National Cohort (GNC) have joined the programme in order to harmonise auditing activities on the national level.

### 3.5 Ring Trials

Interlaboratory comparisons are highly desirable to harmonise and improve preanalytical processes, and to identify factors which may have an impact on sample quality [4]. To provide a consistently high quality of individual samples, a ring trial concept for different bioresources with the focus on continuous improvement of processes has been established within GBA. During ring trials,

reference samples are processed, nucleic acids are isolated and analysed according to the standard operating procedures at the different biobanks. Improvements in handling were discussed in a cross-project exchange with the GBA working group for education of the technical biobank staff. Furthermore, in a pilot ring trial across GBA biobanks, metabolic quality indicators are being validated to become part of the standard quality control procedure for samples before they are stored. The indication whether samples have been tested accordingly will in future be included in GBA Sample Locator [5] as a quality criteria.

### 3.6 Satisfaction Survey for Biobank Users

To support biobanks in meeting their users' needs, GBN designed an online questionnaire. The results of this survey enable biobanks to adapt internal processes to the needs of their customers to ensure sustainability. The survey is conducted once a year. There are two options to perform the survey: (i) Using the online tool provided by GBN or (ii) applying a biobank own survey tool, which rely, however, covers the same questions and, potenially, some further questions regarding the respective biobank. Irrespective of the option used, GBN performs a comprehensive evaluation with the data of all biobanks. An indication whether this survey is being conducted by a biobank will also be included in the Sample Locator.

## 4 Challenges and Opportunities

The belief that the main purpose of biobanks is the collection and storage of biosamples is outdated. Modern biobanks should do as much as possible to ensure that their collected and stored samples are also used for biomedical research. If it comes to the use of samples and data from various biobanks harmonisation of (heterogeneous) data sets and samples quality is an additional aspect [6]. Here, biobank networks such as the GBA are indispensable to ensure harmonised standards in order to achieve reliable and comparable sample and data quality and to enhance the visibility of their sample collections.

In addition, it is essential that biosamples are not only linked to medical information but also to research data to enhance their value and to avoid repetition of data generation [7]. Therefore, data and material transfer agreements should be in place to allow the return of research data based on harmonised metadata according to the FAIR principles [8]. This would ensure a much more sustainable use of sample and research data.

## 5 Outlook

The GBN quality concept for biobanks is a real added value for the biobanking community in Germany. An awareness for quality is built and biobank staff

received targeted training. More biobanks will become GBA partners and contribute their experiences and knowledge as well as benefit from the developed products and the exchange within the community.

Establishing harmonised processes on a national and, ultimately, European level is the aim pursued in cooperation with BBMRI-ERIC. This is especially relevant regarding the EU regulation for in vitro diagnostics, which came into force in 2017 – the transition period will end in 2022. The demand for well-characterised human samples will consequently rise and IVD device manufacturers might lay down additional requirements regarding sample quality.

Sample quality is as well a basic requisite for data quality: the collection, evaluation and use of data will become the most important field of competence for biobanks – in some parts of the world that might already be the case. High-resolution digitalisation of tissue sections for virtual microscopy is becoming accessible for an increasing number of biobanks. Machine learning of many histological patterns linked to the medical data generate algorithms enabling pattern recognition for prognostic and potentially therapeutic purposes [9]. In less than ten years from now, huge data pools and algorithms for pattern recognition will be available in pathology which – in combination with molecular alterations – will allow the precise identification of most effective drugs for most successful therapies.

Key for success are official standards which ensure compatibility and interoperability. In addition, samples and data sharing without unnecessary restrictions is mandatory. This might – at least to some extent – need cultural change. Large infrastructures such as BBMRI-ERIC and its national nodes play a central role in this respect. Back to the future: The year 2029. Obstacles such as quality variations, missing data and insufficient interoperability have been overcome. Biobanks play a major role as sources of high quality biospecimens and comprehensive data of all kind. To accomplish this, biobanks have been embedded in well-developed research infrastructures: the future has already started!

## References

1. BBMRI-ERIC - Biobanking and BioMolecular Resources Research Infrastructure - European Research Infrastructure Consortium. www.bbmri-eric.eu
2. German Biobank Node (GBN): bbmri.de
3. Schmitt, S., et al.: German Biobank Node: Manual for Quality Management in Biobanking - English Version. https://doi.org/10.5281/zenodo.3697691
4. Herpel, E., Schmitt, S., Kiehntopf, M.: Quality of biomaterials in liquid- and tissue-biobanking. Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz **59**(3), 325–335 (2015). https://doi.org/10.1007/s00103-015-2294-3
5. Cooperation between German Biobank Node, German Cancer Research Center in Heidelberg, Biobanking and Biomolecular Resources Research Infrastructure - European Research Infrastructure Consortium. https://samplelocator.bbmri.de/
6. Huppertz, B., Holzinger, A.: Biobanks – a source of large biological data sets: open problems and future challenges. In: Holzinger, A., Jurisica, I. (eds.) Interactive Knowledge Discovery and Data Mining in Biomedical Informatics. LNCS,

vol. 8401, pp. 317–330. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-43968-5_18
7. Müller, H., et al.: State-of-the-art and future challenges in the integration of biobank catalogues. In: Holzinger, A., Röcker, C., Ziefle, M. (eds.) Smart Health. LNCS, vol. 8700, pp. 261–273. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16226-3_11
8. Wilkinson, M.D.: The FAIR guiding principles for scientific data management and stewardship. Sci. Data **3**, 160018 (2016). https://doi.org/10.1038/sdata.2016.18
9. Courdray, N., et al.: Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. Nat. Med. **24**(10), 1559–1567 (2018). https://doi.org/10.1038/s41591-018-0177-5