

# Biobanks in the Era of Digital Medicine

Gunnar Jacobs<sup>1</sup>, Andreas Wolf<sup>2</sup>, Michael Krawczak<sup>2</sup> and Wolfgang Lieb<sup>1</sup>

Digitalization is currently permeating virtually all sectors of modern societies, including biomedical research and medical care. At the same time, biobanks engaged in the long-term storage of biological samples that are fit for purpose have become key drivers in both fields. The present article highlights some of the challenges and opportunities that biobanking is facing in the current proverbial “era of digitalization.”

The ability to store tissue and liquid biosamples that are fit for purpose over a long period of time is a prerequisite for forward-looking biomedical research. First and foremost, biobanking ensures that these samples can be reliably analyzed with up-to-date (for example, “OMICS”) technologies long after the time of their collection. In addition, biobanking allows baseline information about the sample donors to be related to other, prospectively ascertained health parameters (e.g., from population-based cohort studies) or long-term clinical outcomes (e.g., from electronic health records (EHRs) or from self-reports). Therefore, biobanking with strict quality standards is an essential requirement, not only for the development of new diagnostic and prognostic markers, but also for gaining an improved pathophysiological understanding of disease development. At the same time, our societies are currently undergoing comprehensive digitalization, which will likely confront biobanks with both new opportunities and new challenges.

## CONTINUOUS DOCUMENTATION OF THE LIFE CYCLE OF A BIOSAMPLE

Digital information technologies offer the possibility to track the entire life cycle of a biosample, from its original collection and preanalytical handling via the intermediate storage conditions, including freeze–thaw cycles, to its ultimate scientific use. All this information needs to be documented carefully in order to be able to judge whether the quality of a given biosample allows it to be sensibly included for a particular type of research.

Hand-held scanners can be used to initiate the above-mentioned tracking of a biosample already at the bedside. Bar-coded collection tubes and wristbands of patients allow safe and reliable documentation of both the donor identity and the exact time of collection. All subsequent steps of collection, transportation, preparation, and

storage can (and should) be documented as well, preferably using time stamps. Ambient conditions during transport, processing, and storage can be monitored by temperature-logging systems. Combining the acquired data in an integrated “Biomaterial Information and Management System (BIMS),” a comprehensive digital history of individual biosamples can be generated and used in research to benefit both efficiency and reproducibility.

An easy way to document the preanalytical handling of biosamples in standardized fashion is by means of the so-called “Standard PREanalytical Code (SPREC),” which was first devised in 2009 and later developed further by a dedicated International Society for Biological and Environmental Repositories (ISBER) working group.<sup>1</sup> The SPREC has seven elements that facilitate meaningful comparison of quality across different biomaterial collections.

Over and above proper documentation, interest has also arisen in the availability of biomarkers to reflect current biosample quality.<sup>2</sup> In fact, measurement of particularly sensitive and unstable metabolites already provides a means to retrospectively assess the preanalytical handling and long-term storage conditions of individual samples.<sup>2</sup> For example, taking into account the plasma concentrations of ascorbic acid and lactic acid, Trezzi *et al.* proposed a so-called LacaScore as an indicator of whether a given blood sample can be used for metabolomic analyses.<sup>3</sup>

## ONLINE APPLICATION PROCEDURE, VISIBILITY, AND TRACKING OF USAGE OF BIOSAMPLES

It has become widely accepted that the use of biosamples and associated data for scientific research requires structured application procedures, involving both the local ethical review boards and dedicated use and access committees (UAC). Most major biobanks and cohort studies that maintain biomaterial collections (e.g., UK Biobank, Framingham Heart Study, German National Cohort) already rely on “paper-free” online application procedures for their biosamples and data. Digitalization allows detailed tracking of these application and review processes as well (e.g., dates of submission, evaluation and decision, UAC ruling, dates and conditions of data and sample release). Comprehensive monitoring undoubtedly serves to improve both the efficiency and the transparency of the use and access process. Moreover, it allows biobanks to track how and how often their samples and data are

<sup>1</sup>Institute of Epidemiology, University of Kiel, and PopGen Biobank, University Hospital Schleswig-Holstein Campus Kiel, Kiel, Germany; <sup>2</sup>Institute of Medical Informatics and Statistics, University of Kiel, Kiel, Germany. Correspondence: Wolfgang Lieb (wolfgang.lieb@epi.uni-kiel.de)

Received 7 November 2017; accepted 30 November 2017; advance online publication 29 December 2017. doi:10.1002/cpt.968

being used for research. This is an important success criterion for biobanks because the scientific and socioeconomic utility of a biobank is directly linked to its usage intensity. Digitalization also allows detailed information about biosamples available for research to be shared easily with a wider community, and searchability via publicly accessible online platforms is yet another guarantor for the efficient usage of these precious resources (for a commercial solution see, for example, <http://www.ispecimen.com/>). Finally, digitalization helps to improve the visibility of biobanks, for example, through their own website traffic and social media presence.

### COMBINING BIMS WITH OTHER DIGITAL DATA SOURCES

Digitalization will greatly ease the integration of biosample-derived data (e.g., OMICs data) with the wide spectrum of phenotypic data obtained in other dedicated research settings, or extracted from EHRs, or provided by health insurers or patients themselves.<sup>4</sup> However, a number of challenges have to be met in this context. Apparently, sample-related data must be in digital format in the first place to allow storage in the BIMS or in separate databases connected to, and documented by, the BIMS. Second, linkage of sample-related data to other donor-related data (e.g., from EHRs) requires appropriate identity management with due account taken of the donor consent as well. Third, data heterogeneity and quality issues need to be addressed adequately. EHR data, for example, may have to be structured by natural language processing before sensible inclusion in a research database. Finally, most legislation requires research data to be deidentified before storage, including the obfuscation or removal of personal information hidden in the actual data themselves.

Moreover, the wide availability of digital information will promote the generation and use in medical research of metadata, including, for example, the types of assay or device used for a particular analysis or measurement. Over and above this immediate informational value, metadata repositories can also be used to develop common data ontologies for biobanking or to harmonize laboratory measurement conditions and units. All biomaterial-related data sources should be connected to a core BIMS using standard data exchange formats, not only to support and facilitate local biomedical research, but also to enhance interoperability between different research institutions.<sup>5</sup>

### PRIVACY AND DATA PROTECTION

Advancing digitalization in biobanking raises various privacy and data protection issues. First and foremost, it must be ensured that the management and use of comprehensive, biosample-related data is embedded into appropriate organizational information technology infrastructures that comply with the country-specific legal and ethical frameworks. Digitalization also allows the involvement of independent, trusted third parties into the management of donor identity and consent management, thereby providing an additional safeguard against the misuse of personal biomedical information. A key element of appropriate data management in biobanks should be that clinical data, sample-related data, and identifying data are physically stored in separate databases under different administrative power and using different identifiers. However, one particularly tricky legal

and ethical challenge posed by digitalization must not remain unmentioned: due to the ease of generating multiple copies of a given dataset, full deletion of such data will be difficult in cases where a donor withdraws her/his original consent. In fact, quite often such requests for comprehensive deletion will conflict with serious technical, legal, or regulatory constraints, thereby rendering perfect compliance impossible.

In summary, digitalization provides biobanks with great opportunities to improve their performance, including 1) better documentation of the life cycle, quality, and scientific use of biosamples; 2) maintenance of appropriate use and access procedures; and 3) better interoperability with other sources of donor-related data. However, while the collection, integration, and use of multidimensional data from different sources bears great scientific potential, it also poses legal and ethical challenges that need to be addressed appropriately to ensure that biobanking remains an acceptable and widely accepted enabler of biomedical research.

### ACKNOWLEDGMENTS

The authors received funding from the Federal Ministry for Education and Research (01EY1103, 01ZX1306A, 01ZX1606A, 01ER1301/13) related to their biobank activities.

### FUNDING

German Federal Ministry for Education and Research: Michael Krawczak, Wolfgang Lieb: 01EY1103; German Federal Ministry for Education and Research: Michael Krawczak, Wolfgang Lieb: 01ZX1306A; German Federal Ministry for Education and Research: Michael Krawczak, Wolfgang Lieb: 01ZX1606A; German Federal Ministry for Education and Research: Wolfgang Lieb: 01ER1301/13. The authors received funding from the Federal Ministry for Education and Research (01EY1103, 01ZX1306A, 01ZX1606A, 01ER1301/13) related to their biobank activities.

### CONFLICT OF INTEREST

The authors declared no competing interests for this work.

### AUTHOR CONTRIBUTIONS

All authors conjointly wrote the article.

© 2017 The Authors. *Clinical Pharmacology & Therapeutics* published by Wiley Periodicals, Inc. on behalf of American Society for Clinical Pharmacology and Therapeutics

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

1. Lehmann, S. *et al.* Standard preanalytical coding for biospecimens: review and implementation of the Sample PREanalytical Code (SPREC). *Biopreserv. Biobank* **10**, 366–374 (2012).
2. Betsou, F. *et al.* Assays for qualification and quality stratification of clinical biospecimens used in research: a technical report from the ISBER Biospecimen Science Working Group. *Biopreserv. Biobank* **14**, 398–409 (2016).
3. Trezzi, J.P. *et al.* LacaScore: a novel plasma sample quality control tool based on ascorbic acid and lactic acid levels. *Metabolomics* **12**, 96 (2016).
4. Antman, E.M. *et al.* Acquisition, analysis, and sharing of data in 2015 and beyond: a survey of the landscape: a conference report from the American Heart Association Data Summit 2015. *J. Am. Heart Assoc.* **4**, (2015).
5. Kiehn, M. & Krawczak, M. Biobanking and international interoperability: samples. *Hum. Genet.* **130**, 369–376 (2011).