



ELSEVIER

Contents lists available at ScienceDirect

New BIOTECHNOLOGY

journal homepage: www.elsevier.com/locate/nbt

Transnational access to large prospective cohorts in Europe: Current trends and unmet needs

Birgit A. Simell^{a,*}, Outi M. Törnwall^b, Iiro Hämäläinen^{a,c}, H.-Erich Wichmann^{d,e,f}, Gabriele Anton^{d,f}, Paul Brennan^g, Laurene Bouvard^g, Nadia Slimani^g, Aurelie Moskal^g, Marc Gunter^{g,h}, Kurt Zatloukalⁱ, Joel T. Minionⁱ, Sirpa Soini^c, Michaela T. Mayrhofer^b, Madeleine J. Murtagh^j, Gert-Jan van Ommen^k, Mattias Johansson^g, Markus Perola^{a,c,l}, for the BBMRI-LPC Consortium (FP7 GA no. 313010)

^a Institute for Molecular Medicine Finland FIMM, University of Helsinki, Biomedicum 2U, Tukholmankatu 8, 00290, Helsinki, Finland

^b Biobanking and Biomolecular Resources Research Infrastructure, European Research Infrastructure Consortium (BBMRI-ERIC), Neue Stiftingtalstrasse 2/B/6 8010 Graz, Austria

^c National Institute for Health and Welfare (THL), Biomedicum 1, Haartmaninkatu 8, 00290, Helsinki, Finland

^d Helmholtz Zentrum Munich, Institute of Epidemiology, Ingolstädter Landstraße 1, D-85764, Neuherberg, Germany

^e Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology, Ludwig Maximilians University, Geschwister-Scholl-Platz 1, 80539, Munich, Germany

^f Institute of Medical Statistics and Epidemiology, Technical University Munich, Arcisstraße 21, 80333, Munich, Germany

^g International Agency for Research on Cancer (IARC), 150 Cours Albert Thomas, 69372, Lyon CEDEX 08, France

^h Imperial College London, South Kensington Campus, London, SW7 2AZ, United Kingdom

ⁱ Medical University of Graz, Auenbruggerpl. 2, 8036, Graz, Austria

^j Newcastle University, NE1 7RU, United Kingdom

^k Leiden University Medical Center, Albinusdreef 2, 2333 ZA, Leiden, the Netherlands

^l University of Tartu, Estonian Genome Center, Riia 23b, 51010, Tartu, Estonia

ARTICLE INFO

Keywords:

Biobanks

Transnational access

Cohort studies

GDPR

Collaborative research

ABSTRACT

Biobank samples and data from studies of large prospective cohorts (LPC) represent an invaluable resource for health research. Efficient sharing and pooling of samples and data is a central pre-requisite for new advances in biomedical science. This requirement, however, is not compatible with the present scattered and traditional access governance structures, where legal and ethical frameworks often form an obstacle for effective sharing. Moreover, the EU General Data Protection Regulation (GDPR) is demanding increasingly rigorous administration from all those organisations processing personal data. The BBMRI-LPC project (Biobanking and Biomolecular Research Infrastructure — Large Prospective Cohorts) assembled 21 LPCs from 10 countries and two EU-wide multinational cohort networks with a key objective to promote collaborative innovative transnational research proposed by external researchers on the broad field of common chronic diseases, and analyze the gaps and needs involved. BBMRI-LPC organized three scientific calls to offer European investigators an opportunity to gain free of charge transnational access to research material available in the participating cohorts. A total of 11 high-quality research proposals involving multiple prospective cohorts were granted, and the access process in the individual projects carefully monitored. Divergent access governance structures, complex legal and ethical frameworks and heterogeneous procedures were identified as currently constituting substantial obstacles for sample and data transfer in Europe. To optimize the scientific value and use of these research resources, practical solutions for more streamlined access governance in collaborative projects are urgently needed. A number of infrastructure developments could be made to improve time-efficiency in access provision.

Abbreviations: BBMRI, biobanking and biomolecular research infrastructure; DTA, data transfer agreement; EEA, European economic area; ELSI, ethical, legal and social implications; ERIC, European research infrastructure; GDPR, general data protection regulation; LPC, large prospective cohort; MTA, material transfer agreement; PPP, public-private partnership

* Corresponding author at: FIMM, University of Helsinki, Biomedicum 2U, Tukholmankatu 8 (PO Box 20), 00290, Helsinki, Finland.

E-mail address: birgit.simell@helsinki.fi (B.A. Simell).

<https://doi.org/10.1016/j.nbt.2018.10.001>

Received 3 September 2018; Received in revised form 15 October 2018; Accepted 15 October 2018

Available online 17 October 2018

1871-6784/ © 2018 Elsevier B.V. All rights reserved.

1. Introduction

Prospective large population-based cohort studies (i.e. longitudinal studies on a collection of biological, clinical and other data from large numbers of individuals with shared characteristics) have a fundamental role in identifying etiological risk factors for common chronic diseases [1–4]. In contrast to other study designs (e.g. case-control studies), prospective studies are not affected by recall or other biases, as the risk factors are measured prior to the onset of the targeted disease outcomes [5]. A high number of large prospective cohort (LPC) studies have been carried out across Europe over the last decades. These cohorts constitute a mixture of ‘*mature cohorts*’ with an extensive follow-up of several decades, and ‘*contemporary cohorts*’ with more recent and often enriched exposure information. Some of these cohorts have recruited hundreds of thousands of study participants (e.g. UK Biobank and the EPIC cohort), representing major investments of public funding, as well as time and effort from study participants and researchers. With their clinical samples and extensive exposure information stored in biobanks (biorepositories accepting, processing, storing and distributing biospecimens and associated data for use in research and clinical care), the LPC studies represent an invaluable resource for health research [4,6].

To enable analysis of diseases in a reliable manner, prospective cohorts need to be large enough to accrue sufficiently high numbers of incident disease cases during the follow-up. Consequently, studying rare or moderately common endpoints in a single prospective cohort is often challenging. In many cases, pooling resources across multiple cohorts from different countries is required to assemble sufficient sample size for answering important research questions. Consortium-based cohort studies are instrumental to study less common disease endpoints, highlighting the importance of collaborative cohort research in Europe. These multinational research projects, however, are often challenged by sample and data sharing difficulties due to limited resources and heterogeneous, country- and cohort-specific access procedures.

A major logistical hurdle for effective transnational sharing is the

absence of centralized and uniform access governance [7]. Multiple time-consuming processes specified in the national legislations and local rules have to be complied with, which may markedly increase the time needed before the actual research can commence. Heterogeneity in access procedures is not only observed across regions (e.g. the EU), but even between different cohorts/biobanks within a single country [8]. Regrettably, the burden of current fragmentation falls on the scientist and the laboriousness of transnational access procedures tend to impede the actual science [9,10].

While streamlining access governance systems is not a straightforward task, there are already a number of initiatives ongoing to rationalize and harmonize access policies and procedures across Europe. There is a growing international consensus on the need for improved access to optimize the long-term value of available sample and data collections and to exploit their full potential for health discovery and validation [7]. The pan-European BBMRI-ERIC (Biobanking and Biomolecular Resources Research Infrastructure - European Research Infrastructure Consortium) may prove a promising platform for increased access harmonization, centralization and e-governance [11]. Another way to promote more harmonized and streamlined access solutions is an international Charter that outlines guiding principles for international sample and data sharing, accompanied by a general model for a Material and Data Transfer Agreement (MTA/DTA) [7].

In this paper, we report and discuss the specific bottlenecks identified in the transnational access process identified during the four-year (2013–2017) EU infrastructure project BBMRI-LPC. Furthermore, we provide a set of recommendations to improve the efficacy of access provision to the cohort research resources in Europe. This information is envisaged to benefit the entire biobanking and medical research community, as well as to promote more harmonized and accessible biobanks.

2. BBMRI-LPC project

From 2013–2017, the BBMRI-LPC project (EU FP7 GA no. 313010)

Table 1
List of the European LPCs participating in the BBMRI-LPC project.

Cohort name	Study design	Country	Target number of participants	Website
Cohort of Norway (CONOR)	Cohort study	NO	200,000	https://www.fhi.no/studier/cohort-of-norway/
Constances	Cohort study	FR	200,000	http://www.constances.fr/
Danish National Birth Cohort (DNBC)	Cohort study	DK	100,000	http://www.nationalbiobank.dk/
Estonian Genome Project (EGP)	Cohort study, Other	EE	51,535	https://www.geenivaramu.ee/en
European Prospective Investigation into Cancer and Nutrition (EPIC network)	Cohort study	EU-wide	521,468	http://epic.iarc.fr/
Epidemiology of Health (EpiHealth)	Cohort study	SE	300,000	https://www.epihealth.se/
National FINRISK Study	Cohort study	FI	38,700	http://www.nationalbiobanks.fi/index.php/studies2/7-finrisk
Gazel	Cohort study	FR	20,625	http://www.gazel.inserm.fr
Genomes for Life. Cohort Study of the Genomes of Catalonia (GCAT)	Cohort study	ES	50,000	http://www.gcatbiobank.org/
Health 2000	Cohort study	FI	8,028	https://thl.fi/en/web/thl-biobank/for-researchers/sample-collections/health-2000-and-2011-surveys
Nord-Trøndelag Health Study (HUNT)	Cohort study	NO	125,000	https://www.ntnu.edu/hunt
Janus Serum Bank	Cohort study	NO	318,628	https://www.krefregisteret.no/en/Research/Janus-Serum-Bank/
Cooperative health research in the region of Augsburg (KORA)	Cohort study	DE	18,000	https://www.helmholtz-muenchen.de/kora
LifeGene	Cohort study	SE	200,000	https://www.lifeGene.se
Lifelines Cohort Study	Cohort study	NL	165,000	https://www.lifelines.nl/researcher/biobank-lifelines
Norwegian Mother and Child Cohort Study (MoBA)	Case-control study, Cohort study	NO	284,000	https://www.fhi.no/en/studies/moba/
MONICA Risk, Genetics, Archiving and Monograph (MORGAM network)	Other	EU-wide	128,874	https://thl.fi/morgam/
Netherlands Cohort Study (NLCS)	Cohort study	NL	120,852	https://epi.grants.cancer.gov/Consortia/members/nlcs.html
Rotterdam study	Cohort study	NL	15,000	http://www.epib.nl/research/ergo.htm
TwinGene	Cohort study	SE	12,600	https://ki.se/en/research/the-swedish-twin-registry
United Kingdom Biobank (UK Biobank)	Cohort study	UK	500,000	https://www.ukbiobank.ac.uk/
		Total	3,378,310	

Table 2

The access research projects selected in the three BBMRI-LPC scientific calls. A list of scientific publications resulting from the access projects is available at <http://www.bbMRI-lpc.org/access>.

Scientific call	Title of the research project (PI, principal investigator)	Country of origin
1	Early biomarkers for pancreatic cancer (Prof. Gert-Jan van Ommen)	NL
1	Metabolic pathways of kidney cancer (Dr. Mattias Johansson)	FR
1	Markers of imminent myocardial infarction (Prof. Johan Sundström)	SE
1	Exposure to brominated disinfection by-products and prospective metabolic alteration (Dr. Konstantinos Makris)	CY
2	Biomarkers of human papillomavirus (HPV) infection and risk of two increasing cancers (Dr. Tim Waterboer)	DE
2	Metabolic Profiling and Colorectal Cancer Risk (Dr. Marc Gunter)	UK
2	Exploring the genetics of irritable bowel syndrome (Prof. Mauro D'Amato)	ES
2	Biomarkers of ischemic stroke in prospective cohorts (Dr. Christian Gieger)	DE
3	Identification of biomarkers for gallbladder cancer risk prediction (Prof. Bermejo Lorenzo)	DE
3	The influence of macronutrient dietary patterns on pregnancy (Prof. Janet Cade)	UK
3	Unravelling the role of Bile Pigments in Colorectal Cancer (Prof. Karl-Heinz Wagner)	AT

assembled 21 LPC studies from 10 countries and two EU-wide multinational cohort networks (Table 1) with an objective to promote collaborative transnational research on prospective cohorts and analyze the gaps and needs involved. Procedures for bringing together samples and data from participating cohorts were developed and three scientific calls for transnational research proposals using multiple LPCs were organized. For guidance to aspiring users and submission of study proposals, a specific BBMRI-LPC access web portal was launched (<http://bbMRI-lpc.iarc.fr/mica/>) providing information about the support offered, a cohort catalogue of the available resources and instructions for on-line proposal submission. An independent, scientific Review Committee constituted of seven BBMRI-LPC members (including cohort representatives) was responsible for evaluation of access project proposals based on the specific pre-defined criteria. Proposals were ranked according to their average sum of scores across all categories. In total, 26 research proposals were received, of which 11 projects were awarded support (Table 2). Successful applicants received a template letter for approaching the cohorts. The BBMRI-LPC access team facilitated the initial contact and provided the users with advice on the subsequent steps of the access process. The costs of accessing data and samples of the selected cohorts, as well as generation of specific types of omics data, were funded by the BBMRI-LPC grant, whereas the downstream analyses had to be covered by other funds of the applicants. The access process in the individual projects was closely monitored by the BBMRI-LPC access team.

In parallel, an interview study was conducted to understand the barriers and facilitators for sample and data sharing through the experiences in BBMRI-LPC. In the interviews, experiences and perspectives of individuals involved in BBMRI-LPC as well as other professionals or observers in the field were explored. The aim of the interviews was to capture trends and strategies for sample and data sharing, as well as to collect feedback from 'the BBMRI-LPC customers', including the research project applicants and the biobank personnel. The results of the interview study will be published separately (M. Murtagh et al.).

3. Trends and needs in the current access procedures

Any researcher accessing samples or data in a cohort or biobank typically needs to complete three general steps of the access process: (1) requesting the access, (2) obtaining the necessary study approvals, and (3) executing the access. Below, we describe the current bottlenecks for each of these steps as revealed by close follow-up of transnational research projects in the BBMRI-LPC project.

3.1. Requesting the access

3.1.1. Cataloguing of samples and data

Evaluating which individual cohorts are suitable to answer a specific research question is rarely straightforward, as most cohorts do not

provide detailed information on their research resources in a harmonized fashion. To facilitate this process for the BBMRI-LPC-supported access research projects, a specific access web portal (<http://bbMRI-lpc.iarc.fr/mica/>) was launched where the most relevant information on research resources was made easily available in a searchable form. In addition, data in the BBMRI-LPC cohort catalogue, a sub-catalogue of the BBMRI-ERIC catalogue of European biobanks [12], was updated by inviting 21 participating cohorts to take part in a harmonized exposure and clinical endpoint data inventory (<http://www.bbMRI-lpc-biobanks.eu/>). Information on epidemiological exposure data (i.e. any factor, which may be associated with an outcome of interest) was received for 13 cohorts, whereas eight cohorts were unable to provide all relevant information mostly due to insufficient financial resources to support this activity. Information on clinical endpoints (i.e. primary outcomes, such as diseases or other measurable events, measured in a clinical trial) were received for 9 cohorts. For another 12 cohorts, the endpoint data were not available at the time of request.

3.1.2. Centralized support

For researchers formerly unfamiliar with European cohorts, information provided in the cohort catalogue itself was in many occasions not sufficient to allow preparation of competitive research proposals. Supplementary support by the BBMRI-LPC access team was often required to guide investigators in defining the requirements of samples and data for their research project, the specific structure and set-up of individual cohorts, as well as how to approach the cohorts.

3.1.3. Identification of a correct contact person for the access correspondence

In some cases, identification of the correct contact person for access correspondence at the cohort took a surprisingly long time. This caused unnecessary delay in initiating the actual access procedures. It is worth noting that while the principal investigator of the original cohort study is basically the ideal contact person for all scientific matters regarding the cohort, they are generally not the most suitable point of contact for assisting with the various administrative and technical practicalities related to access.

3.2. Obtaining the necessary study approvals

3.2.1. Heterogeneous access governance

Some contemporary European cohorts have adopted a service-oriented access governance structure with minor scientific involvement required in the downstream research. Such cohorts tend to have strong research support to handle the relevant study administrative tasks. Other cohorts, most of which can be considered more mature, often require the research to be conducted in close collaboration with a local scientist, who would also be in charge of handling many of the local study administrative duties. Whilst the role of a local scientist is beneficial in ensuring that the idiosyncrasies of individual cohorts are

Table 3

The bottlenecks identified and recommendations on measures to improve the transnational access provision to European cohorts/biobanks.

Bottleneck	Recommendation(s)	Category
REQUESTING THE ACCESS		
Cataloguing of samples and data	Local investments for cataloguing the cohort resources in a standardized fashion Up to date maintenance of the catalogue information	Financial resources
Centralized access support	A dedicated contact person appointed for access correspondence at the cohort (access manager)	Access governance, personnel resources
GETTING THE ACCESS APPROVED		
Heterogeneous access governance structures	Harmonization of the access policies nationally and internationally An access research proposal template pre-agreed by the cohorts	Access governance
Time-consuming local approval processes	Information on the local decision bodies and their requirements available Strong administrative organization of the local decision bodies	Access governance
EXECUTING THE ACCESS		
MTA/DTA negotiations	Standardized MTA/DTA or a collection of standardized building blocks. Initiation of the MTA/DTA review process early upon project approval Importance of agreeing on the pre-determined time frames	Access governance
Delays in organizing and shipping of study samples	Local investments in biobanks' personnel and sample management automation.	Financial resources
Data harmonization	A cohort-wise one-off harmonization of common study variables	Financial resources

appropriately accommodated in the study setup and conduct, this approach is often rate-limiting as the local scientist may have limited time to handle such tasks. Regardless of governance structure, individual cohorts each have their own procedures and rules for organizing access to their samples and data, requiring individual users who seek access to multiple cohorts to comply with multiple, and often strongly divergent, access procedures. The scope and validity of informed consent from the research participants is often subject to evaluation and may limit the secondary uses and data sharing, even if the actual access procedures were smooth. The new EU General Data Protection Regulation (GDPR) became effective on May 25, 2018, and imposes reassessment of the validity of existing informed consents, if they are used as a lawful basis for research.

3.2.2. Time-consuming local approval procedures

One of the most time-consuming tasks in the access process is obtaining the necessary local scientific and ethical approvals. Most cohorts have a local scientific access committee or an equivalent decision body for evaluating new access requests. In many cases information about these cohort-specific committees and their requirements only became known following the initial access request. As meetings of the local committees are typically organized at monthly or bi-monthly intervals, gaining approvals from each of the multiple cohorts caused inevitable and varying delays. On some occasions, applicants were also asked to modify their proposal or to provide complementary information, which led to additional round(s) of committee review(s) before the final decision was made.

Whilst some selected cohorts have provided overarching ethical approval for future use of their data and samples, most require local or national ethics committees (or similar body) to approve any proposed research wherein the cohort's data and/or samples would be accessed. Some national laws also require a review from the data protection authorities. The local ethics committees meet at regular intervals, but rarely in conjunction with the local access committees, which tends to generate an additional delay in the approval process. In the cohorts with a very streamlined process, the time taken by the local access and ethics committees to grant approval was approximately three to four months, while in the worst case the approval process took more than a year.

3.3. Executing the access

3.3.1. Material Transfer Agreement (MTA) negotiations

Following approvals from the local access and ethics committees, establishing the necessary MTAs that govern use of the cohort data and samples proved to be a complicated administrative effort, typically

taking many months up to one year to complete. Initially, BBMRI-LPC developed a uniform MTA template that was offered as an option to be used by the supported research projects. Despite this effort, the MTA negotiations remained restrictively cumbersome. As there are few important differences in laws or regulations at national level which should complicate the transfer of samples or data within the EU, the problem seemed rather to lie at the institutional level, as most cohorts expressed a strong preference for using their own specific MTAs. Transfers of personal data to non-EU or non-European Economic Area (EEA) countries complicate agreements even further due to the EU data protection regime that requires adequate safeguards to be fulfilled by the recipient.

3.3.2. Delays in organizing and shipping of study samples

Whilst the cohorts generally have well-organized routines for retrieving, preparing and shipping samples, in some instances 3–12 months' delay occurred in retrieving the relevant samples and preparing them for transfer. The extended sample retrieval times were interpreted to be largely dependent on the availability of personnel as well as on the extent of automatic sample manipulation implemented in each of the respective cohorts.

3.3.3. Data harmonization

In contrast with sample access, organizing and transferring cohort data was relatively straightforward and further facilitated by the cohort catalogues. Here, however, a more important bottleneck turned out to be the downstream data harmonization. While this activity was not supported under the present BBMRI-LPC project, it needs to be commented on here as an important hurdle in multinational cohort science.

4. Smoother access solutions for the cohort-based research projects

Below, we provide a set of recommendations for addressing the bottlenecks identified in collaborative cohort-based transnational research. The identified challenges and subsequent recommendations are summarized in [Table 3](#).

4.1. Importance of cataloguing the cohort resources

The importance of a searchable database with detailed, standardized information on the available cohort samples and data is evident. Further, to maintain such a catalogue up-to-date as well as to allow enriching the catalogue through generation of new data, sufficient personnel resources along with appropriate IT systems for (semi-) automated updates of the individual cohort information are necessary. In

the long term, great progress would be made by developing an extended version of the cohort catalogue that supports more detailed search requests and allows direct linkage of specific disease cases with information on the availability of relevant samples and data.

4.2. Enhanced access support services

There is a clear need for a more efficient support model to better assist European scientists in navigating through the sample access procedures. To enhance support for access services at the cohort-level, establishing a professional access manager would be a great advantage. A dedicated scientific administrator serving locally as a centralized contact point for all access correspondence and providing guidance to researchers with the pertinent approval procedures and necessary contracts would not only promote scientific collaboration and communication, but also bring greater visibility for a respective cohort or biobank. Smooth access procedures require appropriate resourcing.

4.3. Efforts to harmonize access governance structures and approval processes

In the cohorts, the ethical, legal, administrative and logistic decisions are typically segmented across specific committees. On many occasions, however, information on these decision bodies and their requirements is not sufficiently clear or publicly available. In addition to publishing the cohort-specific requirements, provision of a pre-agreed, uniform proposal template for new access requests would greatly streamline the overall approval process of projects involving multiple cohorts.

Despite the time taken by assessment of the local access committees, they are still considered as an inevitable step in the approval process. Whilst it would be appealing, for the sake of efficiency, to replace the local access committees with one single national body dealing with all transnational cohort-based research, clearly only a few cohorts would be willing (or even allowed) to relinquish their governance framework in favor of a centralized protocol. Such a body would also require major trust from its constituents to have sufficient knowledge of the individual cohorts. In the short term, it is therefore crucial that local access committees have a strong administrative organization with clear and streamlined routines to work as efficiently and transparently as possible.

In contrast, as the ethical principles, laws and data security regulations within European countries are relatively uniform, having a common independent ethics review board at the European level (such as the 'BBMRI-ERIC Ethics Check' available at <http://www.bbmri-eric.eu/BBMRI-ERIC/ethics-check/>), for cohort-based research projects could considerably enhance the approval processes, although this would require extensive 'buy-in' from individual cohorts and governing institutes to be useful. For now, we strongly support one national ethics review per country for the consortium-based research projects.

4.4. Standardized MTAs

The ethical and legal framework of European institutes has tightened over the last decade with important repercussions for establishing the agreements that govern national and transnational transfer of research material. In practical terms, this means that many institutions have developed increasingly complex MTAs that typically require case-by-case negotiations involving two institutions (often from different countries) to agree on complicated legalese prior to any new sample or data transfer. As much as the scientists in nationally and internationally funded research are nowadays held to time-limited deliverables and milestones, it is crucial that the legal departments involved in negotiating the necessary MTAs provide timely support. We emphasize the importance of pre-determined timeframes, as the obscurity of the schedule may cause unnecessary friction between the parties involved.

By agreeing the timeframe beforehand, it is possible to avoid juxtaposition between the actors and ensure a timely local dialogue between the scientists and the administrators, improving transparency and accountability from both sides.

Many consortium projects have put a lot of effort into developing standardized MTA templates [7]. However, without including the legal departments in the development phase to have the majority of institutes agreeing on such a template, practical implementation thereof is not realistic. Establishing a truly uniform MTA template (or at least a collection of standardized building blocks) has the potential to revolutionize the rate by which transfer of data and samples can occur in Europe, hence substantially accelerating biomedical discoveries. This is, however, a challenging task that requires close collaboration and mutual agreement of each cohort-associated institution. Failing this, applicants in consortium-based research projects will continually be obliged to establish a number of different MTAs. In any given research project requiring transfer of data and/or samples, we strongly recommend initiating the MTA processes as early as possible, ideally before or as soon as all necessary study approvals are in place.

4.5. Investments into sample management and harmonization

Sufficient and project-independent core funding is necessary to efficiently fulfill access claims and swiftly process the sample requests in the individual cohorts. Whilst for some cohorts a certain gain in efficiency might be achievable through organizational changes, substantial reduction of sample preparation time would typically require increasing the personnel resources involved in sample retrieval and manipulation and/or introducing automation, which would again require additional investments in machinery and robotic sample handling.

In terms of data harmonization, it can often take several months for a skilled computer technician to harmonize a complex study database. To avoid having each consortium re-harmonizing the same cohorts' data for each new project, we recommend individual cohorts to assign resources allowing one-off harmonization of common study variables required in any consortium-based research project [13,14]. As an example of such an effort, in 2013 the US National Cancer Institute supported a comprehensive harmonization effort of a large number of commonly used study variables for cohorts participating in the Diabetes and Cancer Initiative (n = 28). The code book for this harmonization will be made available for other investigators seeking to harmonize data from these cohorts. Also, the previously initiated large collaborative FinnGen project (<https://www.finnngen.fi/en>) employs clinical expert groups to harmonize endpoint data acquired from the national health registries. Adopting this kind of principle would provide one of the most cost-efficient improvements in cohort-based research. As the data evolves over time, smart and appropriate IT-solutions are necessary for keeping the information up-to-date and documenting the data provenance. Also, innovative ways of sharing data are needed. When samples become data via laboratory analyses, it may be feasible to "bring analysis to the data instead of bringing the data to the analysis" [15], rather than go through all the steps described above. It should be noted, however, that by GDPR, having access to data is equal to data transfer, albeit the data download would have been made technically impossible.

5. Conclusions

Sharing samples and data from different study sources is crucial in attaining new biomedical breakthroughs. Divergent access governance structures, complex legal and ethical frameworks and heterogeneous cohort access procedures currently constitute substantial obstacles for sample and data transfer in Europe. In addition, lack of resources in the publicly funded cohorts is an issue. Public-private partnership (PPP) in large projects, such as FinnGen (<https://www.finnngen.fi/en>), a collaborative research project of nine Finnish biobanks and seven pharma

companies, is an example of how to overcome the funding issues that are often related to the large projects, particularly in small countries. On the other hand, PPP naturally creates other layers of complexity which need to be dealt with. While the GDPR (EU 2016/679) will harmonize EU practices on personal data protection at the meta-level, a lot of decision power, particularly on scientific projects, has been left for the member countries. GDPR also sets out rigorous provisions for data processing, administration and privacy impact assessments. The validity of old consent may be at stake, if used as a lawful basis for data processing. This will probably not ease the challenges described above, but rather re-boot the field in many ways that make it necessary for the scientists and ELSI (ethical, legal and social implications) experts to map the situation all over again. Streamlining of access provision processes has the potential to greatly enhance use of the unique and invaluable cohort research resources and to maximize their impact on high-quality health science. This is a mandatory route, as it is the only way to provide health care benefits and life style improvements to the public at large, justifying the efforts of the biobank participants. All in all, if the data and samples are used for the original purpose for which they were collected, as well as in accord with the informed consent for their use, we see that all unnecessary restrictions and artificial boundaries should be curtailed in the interest of public good.

Funding

This work was supported by the European Commission Seventh Framework Programme [grant no. 313010].

Declarations of interest

None.

References

- [1] Muller DC, Murphy N, Johansson M, Ferrari P, Tsilidis KK, Boutron-Ruault M-C, et al. Modifiable causes of premature death in middle-age in Western Europe: results from the EPIC cohort study. *BMC Med* 2016;14:87. <https://doi.org/10.1186/s12916-016-0630-6>.
- [2] Muller DC, Johansson M, Brennan P. Lung cancer risk prediction model incorporating lung function: development and validation in the UK biobank prospective cohort study. *J Clin Oncol* 2017;35(8):861–9. <https://doi.org/10.1200/JCO.2016.69.2467>.
- [3] Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015;12(3):e1001779. <https://doi.org/10.1371/journal.pmed.1001779>.
- [4] Brennan P, Perola M, van Ommen GJ, Riboli E. Chronic disease research in Europe and the need for integrated population cohorts. *Eur J Epidemiol* 2017;32(9):741–9. <https://doi.org/10.1007/s10654-017-0315-2>.
- [5] Everitt BS, Palmer Cr. *Encyclopaedic companion to medical statistics. Encyclopaedic companion to medical statistics*. London: Hodder Arnold; 2005.
- [6] De Souza YG, Greenspan JS. Biobanking past, present and future: responsibilities and benefits. *AIDS* 2013;27(3):303–12. <https://doi.org/10.1097/QAD.0b013e32835c1244>.
- [7] Mascalzoni D, Dove ES, Rubinstein Y, Dawkins HJS, Kole A, McCormack P, et al. International Charter of principles for sharing bio-specimens and data. *Eur J Hum Genet* 2015;23(6):721–8. <https://doi.org/10.1038/ejhg.2014.197>.
- [8] Murtagh MJ, Turner A, Minion JT, Fay M, Burton PR. International data sharing in practice: new technologies meet old governance. *Biopreserv Biobank* 2016;14(3):231–40. <https://doi.org/10.1089/bio.2016.0002>.
- [9] Kaye J, Briceño Moraia L, Mitchell C, Bell J, et al. access governance for biobanks: the case of the BioSHaRE-EU cohorts. *Biopreserv Biobank* 2016;14(3):201–6. <https://doi.org/10.1089/bio.2015.0124>.
- [10] Yuille MM, Feller PI, Georghiou L, Laredo P, Welch EW. Financial sustainability of biobanks: from theory to practice. *Biopreserv Biobank* 2017;15(2):85–92. <https://doi.org/10.1089/bio.2017.29019.mjy>.
- [11] Reichel J, Lind A-S, Hansson MG, Litton J-E. ERIC: a new governance tool for Biobanking. *Eur J Hum Genet* 2014;22(9):1055–7. <https://doi.org/10.1038/ejhg.2014.6>.
- [12] Wichmann HE, Kuhn KA, Waldenberger M, et al. Comprehensive catalog of European biobanks. *Nat Biotechnol* 2011;29(9):795–7. <https://doi.org/10.1038/nbt.1958>.
- [13] Doiron D, Burton P, Marcon Y, et al. Data harmonization and federated analysis of population-based studies: the BioSHaRE project. *Emerg Themes Epidemiol* 2013;10(1):12. <https://doi.org/10.1186/1742-7622-10-12>.
- [14] Fortier I, Raina P, Van den Heuvel ER, et al. Maelstrom Research guidelines for rigorous retrospective data harmonisation. *Int J Epidemiol* 2017;46(1):103–5. <https://doi.org/10.1093/ije/dyw075>.
- [15] Gaye A, Marcon Y, Isaeva J, LaFlamme P, Turner A, et al. DataSHIELD: taking the analysis to the data, not the data to the analyses. *Int J Epidemiol* 2014;43(December (6)):1929–44. <https://doi.org/10.1093/ije/dyu188>.