Samples and data accessibility in research biobanks: an explorative survey

Marco Capocasa, Paolo Anagnostou, Flavio D'Abramo, Giulia Matteucci, Valentina Dominici, Giovanni Destro Bisol, Fabrizio Rufo

Biobanks hold human biological samples and/or data giving a crucial contribution to the progress of biomedical research. However, the effective and efficient exploitation of these resources depends on their accessibility. In fact, making bio-resources promptly accessible to all, can favour collaboration among research groups as well as multidisciplinarity. Although this has become a rather common belief, several laboratories still apply secrecy and withholding of samples and data. In this study we conducted a questionnaire based survey in order to investigate sample and data accessibility in research biobanks operating all over the world. 46 out of the 238 contacted biobanks have decided to participate. Most of them provide permission to access their samples (95.7%) and data (85.4%), but free and unconditioned accessibility seems not to be a common practice. The analysis of the biobanks guidelines regarding the accessibility of their resources reveal the importance of three aspects: (i) request for applicants to explain what they would like to do with the required resources; (ii) the role of funding, public or private, in the establishment of fruitful collaborations between biobanks and research labs; (iii) request of co-authorship in order to give access to their data. These results suggest that economic and academic aspects are involved in determining the extent of sharing of samples and data stored in biobanks. As a second step of this study, we investigated the reasons behind the high diversity of the requirements for accessing to biobanks' resources. The analysis of informative answers suggested that the different modalities of resource accessibility seem to be largely influenced by both social context and legislation of the countries where biobanks operate.

Samples and data accessibility in research biobanks: an explorative survey

Marco Capocasa[1*], Paolo Anagnostou[1,2*], Flavio D'Abramo[3*], Giulia Matteucci[1], Valentina Dominici[1,2], Giovanni Destro Bisol[1,2] and Fabrizio Rufo[1,2]

[1]Istituto Italiano di Antropologia, Rome, Italy
[2]Department of Environmental Biology, Sapienza University of Rome, Rome, Italy
[3]Charité Comprehensive Cancer Center, Berlin, Germany

* These authors have equally contributed to the work

Corresponding author:

Marco Capocasa
Istituto Italiano di Antropologia, c/o Sapienza University of Rome, Piazzale Aldo Moro 5, 00185, Rome, Italy
phone number: 0039 06 49912725
e-mail: marco.capocasa@uniroma1.it

Abstract

Biobanks hold human biological samples and/or data giving a crucial contribution to the progress of biomedical research. However, the effective and efficient exploitation of these resources depends on their accessibility. In fact, making bio-resources promptly accessible to all, can favour collaboration among research groups as well as multidisciplinarity. Although this has become a rather common belief, several laboratories still apply secrecy and withholding of samples and data. In this study we conducted a questionnaire based survey in order to investigate sample and data accessibility in research biobanks operating all over the world. 46 out of the 238 contacted biobanks have decided to participate. Most of them provide permission to access their samples (95.7%) and data (85.4%), but free and unconditioned accessibility seems not to be a common practice. The analysis of the biobanks guidelines regarding the accessibility of their resources reveal the importance of three aspects: (i) request for applicants to explain what they would like to do with the required resources; (ii) the role of funding, public or private, in the establishment of fruitful collaborations between biobanks and research labs; (iii) request of co-authorship in order to give access to their data. These results suggest that economic and academic aspects are involved in determining the extent of sharing of samples and data stored in biobanks. As a second step of this study, we investigated the reasons behind the high diversity of the requirements for accessing to biobanks' resources. The analysis of informative answers suggested that the different modalities of resource accessibility seem to be largely influenced by both social context and legislation of the countries where biobanks operate.

Keywords: open science, data sharing, human subjects, research ethics, biorepository.

Introduction

Biobanks play a crucial role in the biological research involving human subjects, providing a fundamental contribution to the rapid growth of the scientific endeavour. This is well demonstrated, particularly in the past ten years, by the funding effort many countries have carried out for the building up of such infrastructures and the management of the biological resources they store (Kaye, 2011). Biobanks hold human biological samples and/or data to facilitate research over time (Wolf et al., 2012). Their development across the world, together with the parallel advancements of laboratory technologies, have dramatically increased the opportunities of studying collections of bio-specimens (and their related data), with broader perspectives than those possible by small collections maintained by single research groups (Haga and Beskow, 2008). However, in addition to the creation of these unprecedented opportunities, the rapid evolution taking place in the field of biobanking has also put researchers in the condition to face a series of new challenges associated with the huge potential benefits of access to biological resources and the several difficulties for their exploitation.

Once detached from the body of the donor, human biological samples become autonomous materials. However, this is only one of the possible dimensions in which they can be represented. In fact, the development of new approaches of genetic and genomic DNA sequencing is leading to an increasing identification of bio-specimens as source of data useful to the progress of biomedicine. From this point of view, biological samples are thus characterized by both material and an informational nature (Macilotti, 2010). These two faces are mainly differentiated by the fact that data, unlike samples from which they were extracted, may, even after the detachment from the body, provide elements that facilitate the identification of the individual donor. This aspect is controversial and it has been particularly debated in light of the results of several recent studies that have warned of the risk of re-identification of donors through a cross-analysis of

genetic data and genealogical metadata available online (e.g. see Bohannon, 2013; Rodriguez et al., 2013). The differentiation between biological materials and data is also substantial from the legal point of view. In fact, while the treatment of the former has to deal with property rights, the issues related to the latter are mainly linked to the fact that it is the expression of the biological identity of the subject. Furthermore, marked differences in strategies put in place to regulate the relationship between these two dimensions, result to be strictly associated with the type of legal system established in the countries where the biobanks operate. This is particularly evident comparing procedures adopted in the United States with those implemented in Europe. In the United States, biological samples are mainly perceived as materials, whereas in Europe we could even considered them as "data carriers" (Macilotti, 2010).

In order to draw a more detailed picture of how biobanks manage their resources, as well as to consider the relationships (and even the contradictions) between the material and the informational spheres of biological samples, we must also take into account the propensity of these institutions to share bio-specimens and data across scientific communities. The first challenge for biobanks consists in finding an equilibrium between the scientific interests of researchers and the expectations of donors. This can be reached by exploiting at best the capabilities and flexibility of current forms of informed consent (Kaye, 2012; Macilotti, 2013; Colledge et al., 2014; D'Abramo, 2015). However, the design of an informed consent able to guarantee the sustainability of resources availability does not solve the issues related with the economic interests usually hidden behind the scientific research. This is the case of several web services selling direct-to-consumer genetic tests and thus accumulating large amounts of samples and data that however, remain unavailable to most research communities and groups (e.g. deCODEme, 23andme, Navigenics; see Knoppers, 2010). Finally, even if biobanks embrace the open science principles, many bioethical issues can emerge as sample and data sharing policies

are different from country to country. In fact, the existence of local legislation ensures compliance with habits and values characterizing the socio-cultural contexts in which biobanks operate (Kaye, 2006; Haga and Beskow, 2008). On the other hand, a widespread and efficient sharing of bio-resources from different countries can only be assured through the achievement of a global consensus on the legislation, the standards and the modalities to be followed. Starting from the preparation of the informed consent, the biobanks' staff must take into account a number of issues when planning prospectively the management of the samples and data. They have to meet the requirements imposed by ethics committees, overcome the difficulties in explaining the future uses of existing samples and put the potential donor in a condition that will allow him to make a really informed decision (Colledge et al., 2014).

Bearing in mind these premises, it cannot be denied that the progress of human biological research largely depends on the openness of resources and scientific knowledge. Epistemic (scientific hypothesis and point of views driving the production of knowledge) and non epistemic (moral, social, political and cultural) values of sharing must be addressed to better understand the importance of supporting the advancement of "open science" in its multifaceted and broad sense. Making bio-resources promptly accessible to all, undoubtedly provide more opportunities for collaboration and encourage multidisciplinarity. Although this has become a rather common belief, several laboratories still apply secrecy and withholding of samples and data (Nelson, 2009; Cadigan et al., 2014; Destro Bisol et al., 2014). The tension raising between the responsibilities researchers have towards the tax paying public and their individual needs (often referring solely to academic and scientific interests modulated through "epistemic values") has its counterpart in the perception of the scientific community regarding biobanks and their services. Milanovic, Pontille and Cambon-Thomsen (2007) have clarified this concept, defining biobanks as "ambiguous entities" that "might be presented as places for archival storage of a cultural

patrimony freely accessible for relevant activities, or as commercial enterprises with lucrative potential". At the same time, participants to biobanks' researches have also raised concerns about the fact that, over particular conditions, private and commercial interests in biobanking may prevail public good and so leading to social tensions (Godard et al., 2010).

The importance to identify solutions to satisfy the needs of both researchers and citizens is well testified by the engagement of a political economic structure such as the *Organisation for Economic Co-operation and Development* (OECD) in supporting open access to publicly funded research products. In fact, in the report *Promoting access to public research data for scientific, economic and social development*, members of an OECD Follow-Up Group (Arzberger et al., 2004) recommend the adoption of open access policies with the aim to exploit the full potential of knowledge and to provide more returns from the public investment. Research involving human subjects are usually conducted following the statements of the Belmont Report, a document of the *National Commission for the Protection of Human Subject in Biomedical and Behavioral Research* defining widely diffused guidelines (http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html). However, this document has been often criticized for its lack of attention to the interaction between researchers and donors (Levine, 1988; Weijer, Goldsand & Emanuel, 1999; Blee and Currier, 2011).

Previous studies conducted on European and U.S. biobanks provided information on the developing trends of biobanking, giving detailed pictures of the type of sample and data stored therein (Hirtzlin et al., 2003; Zika et al., 2011; Henderson et al., 2013). Other studies investigate the opinion of participants and the public about the relationships between sample and data sharing practices and privacy concerns (Kaufman et al., 2009; Lemke et al., 2010). However, to date, only few studies have faced the issue of sample and data sharing behaviour of research biobanks (e.g. see Milanovic, Pontille & Cambon-Thomsen, 2007; Pereira, 2013). The present work aims

to investigate sample and data accessibility in research biobanks operating all over the world by means of a questionnaire based survey. We observed a low rate of free accessibility for both data and biological samples while the requirements for accessing to the non open resources were found to be highly heterogeneous. In order to evaluate the reasons of this heterogeneity, we analysed the relationships between sharing strategies and legal frameworks of the countries in which biobanks operate.

Methods

In this study, we considered the definition of "biobank" as a repository that stores human biological samples, with or without accompanying them with genetic or clinical data (see Haga and Beskow, 2008). Therefore, we have not taken into account neither non-human bio-repositories nor on-line databases. The online survey has been administered to a total of 238 biobanks (see Table 1) operating in Europe (95), America (104), Asia (25), Africa (2) and Oceania (12). The list of biobanks has been acquired through a research using Google, Google Scholar and Pubmed as search engines. The research has been based on the following keywords: biobank", "research biobank" and "human biobank".

The questionnaire was developed to obtain a detailed picture of the sampling activities, the sample and data accessibility criteria and the legal frameworks for their access. The questionnaire consisted of 21 questions (9 closed and 12 open-ended) organized in three sections (see Supplemental File S1). The first section (General information) referred to the name and the place where the biobank operates and other information regarding the funds, the sampling criteria adopted and the type of biological resources stored (sample and/or data). The second section (Biological samples) investigated the sample collection, the ethical requirements and the legal

framework to which the biobank refers to for the management of accessibility to biological samples. The last section (Data) included questions regarding the data collection and the legal framework to which the biobank refers to for regulation of data accessibility. The questionnaire was built using Google Forms (http://www.google.com/forms/about/) and the survey participation was proposed by e-mail sent to each respective biobanks' contact explaining the aims of our research. We sent four sequential invitations to participate (April 18, April 28, May 5 and May 19, 2014), closing the survey at the end of May 2014. As previously stated in the invitation form, the administration of the questionnaire was carried out anonymously. Neither personal information nor biobanks' names were disclosed to others in managing the dataset.

Descriptive statistics of the answers to the closed questions have been performed using Microsoft Excel 2010. Open questions have been analysed following criteria of clarity and informativeness of answers subdividing them in three categories: exhaustive answer (it provides clear and complete explanation of the question); partial answer (it lacks several aspects for a comprehensive description); elusive or no answer (it does not provide any information requested). Furthermore, since many of the answers provided links to external resources (e.g. web links), we also evaluated the exhaustiveness of these documents in order to acquire the information needed to fulfil the questions. When the external references failed to provide clear information due to their difficult retrievability or language other than English, we classified the answer as partial or elusive. Data have been uploaded as online supporting information (Supplemental File S2) and deposited in Zenodo (DOI: 10.5281/zenodo.17098).

Results

General information of the responded biobanks

A total of 46 biobanks operating in four continents replied to our survey invitation: 26 in Europe, 16 in America, 2 in Asia and 2 in Oceania (Table 1). Most of the participant institutions are from United States (30.4%, n=12) followed by the United Kingdom (13.0%, n=6), Italy (10.9%, n=5) and Germany (6.5%, n=3).

More than half of the sampled biobanks are publicly funded (58.7%), whereas the 23.9% and 17.4% make use of private funds or both types of funds, respectively. Interestingly, some different continental situations may be observed. In Europe the rate of institutions that receive only public funds is three times higher than that observed in the American continent (73.1% and 25.0% respectively) whereas the biobanks that make use of both types of funding is not significantly different (19.2% and 18.7%, respectively). The totality of the Asian and Oceanian biobanks analysed here are only publicly funded, but the low number of institutions surveyed (only 4) do not allow any comparison with the other continents.

Looking at sampling criteria used by biobanks to collect bio-specimens, most of them focused their attention only on disease based samples (41.3%) or coupling it with other criteria such as type of tissue (17.4%) or the geographic area where the samples were collected (8.7%). Only seven biobanks focus their attention to criteria other than diseases: three on geography (6.5%), two on types of tissues (4.3%), whereas a population based sample collection is followed by two institutions (4.3%).

Regarding the storing activities, a wide range of biological materials have been collected by the sampled biobanks (e.g. blood, plasma, serum, urine, saliva, nucleic acids, cell lines). Eight institutions store only bio-specimens and operate in the United States (3) , in Europe (3; Italy,

Sweden and United Kingdom) in Asia (2). The remaining 38 biobanks store both biological samples and data (89.1%).

Bio-specimens collection and accessibility: legal and ethical aspects

In the first open question we asked for the ethical requirements followed by the biobank for the sample collection procedures (Question B2; see Figure 1). We mostly focused on the consent obtained from participant (if any) and on approval by a third party (e.g. Ethics Committee, Institutional Review Board). Twenty-two biobanks (47.8%) provided informative answers, referring in all the cases to the consent procedures and often pointing to guidelines, specific local or international laws, and approval by ethics committees or institutional review boards. Open consent seems to be the most utilized manner to involve donors, whereas informed consent results to be less adopted. Some biobanks provide information on privacy issues describing, in most cases, the anonymized character of the personal data. Very few answers stressed the possibility for donors to quit the biobank research (opt-out option). We sorted out 15 answers (32.6%) as semi-informative since they only refer to third parties' responsibility for the sample collection procedures, without spelling any other description regarding the type of consent used (waived or presumed consent included), or else when a reference to specific documents was made (e.g. certifications or laws) this was not easily readable/accessible. Nine answers (19.6%) were not informative or because left blank or referring to vague documents/criteria.

Concerning the strategies of collection and storage of biological samples, we found 24 biobanks (52.2%) that do not accept samples from outside research groups, with a roughly similar percentages in Europe (41.7%) and America (50.0%) continents. On the contrary, 22 biobanks (47.8%) also store biological samples collected by external research groups. Sixteen of them operate in Europe (72.7%), 4 in America (18.2%), 1 in Asia (4.5%) and 1 in Oceania (4.5%). The

majority of them (86.4%), in order to accept samples for storage from external groups, ask from the latter to respect the same ethical requirements adopted by the biobank itself in its sampling procedures. All the biobanks analysed make it possible for researchers to gain access to their bio-specimen collection. Among them, only 2 biobanks [1 European (Estonia) and 1 American (USA)] offer free and unconditioned accessibility to their samples, whereas the remaining 95.7% (44 out of 46) require specific conditions to be satisfied in order to give permission to access their samples. However, our request of specifications regarding the accessibility criteria (Question B4.1; see Figure 1) obtained only 12 informative answers where at least one criterion has been indicated. The analysis of these answers highlights how the sample accessibility seems to be linked to whether the applicants specify their research aims (e.g. studies on a defined disease) and/or the origin of research funds (public, private or both). We also considered as informative those answers indicating that samples are for sale, or when one of the criteria here above listed were specified and readable in external links provided in the answer. Among the answers, 25 were classified as semi-informative. We defined the answers as semi-informative when was indicated that the access is decided by third parties (e.g. IRBs, ethics committees), when a vague criterion was stated (e.g. research project relevance, or researchers working in the public interest), when were indicated specific agreements without a description or not easily readable (i.e. in languages other than English or Italian), or when it was the biobank institution had a person responsible for the access to biological samples. Finally, we sorted out 7 not informative answers that was left blank or referring to agreements or documents not accessible at all.

More than half of the biobanks (54.3%) refer to a specific legal framework for the access to their biological samples. However, only 16 biobanks (34.8%) provided informative answers showing that there are no shared standards but different approaches influenced by social context in which they operate (Question B5.1; see Figure 1). The possibility to gain access to sample seems to

depend mainly from the approval of ethical committees, scientific bodies or bilateral agreements (some biobanks also provided an indication on the model followed, e.g. OECD recommendations or legal contract following laws on customs commerce). However, national laws (e.g., the Italian Garante della Privacy, German Data Protection Laws, Spanish Act 14/2007 and Spanish Royal Decree 1716/2011), Communal international regulation (i.e. the European legal framework), or criteria indicated in international agreements should also be taken into consideration when a request of access to a collection of biological samples is presented. Concerning the possibility of finding documents relative to the legal framework followed by the biobanks, only 31% provided detailed information (Question B5.2; see Figure 1).

Data collection and accessibility: legal and ethical aspects

Most of the surveyed biobanks (41 out of 46; 89.1%) store data extracted from the analysis of their own samples. Differently from what observed for biological samples, 23 out of 41 biobanks (56.1%) also store data produced by external research groups that have used their samples. Among them, a considerable rate (73.9%) require from the external research groups the compliance with the same legal framework they follow.

A slight difference between sample and data sharing propensity has been seen as regards their degree of accessibility. In fact, 3 biobanks [7.3%; 2 Americans (USA) and 1 European (Sweden)] do not allow any access to their data, whereas other 3 biobanks [7.3%; 1 American (USA) and 2 Europeans (France and Italy)] claim to follow a strict open data policy, giving completely free access to their data. However, just like bio-specimens, the majority of biobanks (35 out of 41; 85.4%) allow external research groups to access their data in compliance with certain conditions. We asked them which accessibility criteria they adopt (Question C2.1; see Figure 1). Seven of these biobanks (20%) gave us informative answers, describing codification procedures, providing

reference to specific guidelines, giving access to projects focused on specific groups of diseases, or stating clear criteria (e.g. co-authorship). Other 20 biobanks (57.1%) provide semi-informative answers since they whether referred to (i) third parties authorization for data access (i.e. ethics committees, IRBs, scientific boards); (ii) criteria attached to decision of a single researcher within the biobank institution; (ii) external documents not clearly stating the criteria adopted for data sharing; (iv) other vague criteria (e.g. application for data access through a letter of intent). Finally, among the biobanks giving conditional access, eight (22.9%) provided not informative answers whether because not proper or because the criteria was highly vague (e.g. access given to authorized personnel, access given for research made in public interest).

Similarly to what we observed for bio-specimens, also for data most of biobanks (57.1%) refer to a variety of legal frameworks for their access, depending on the legislation of the country they mostly operate. Among the 35 biobanks (76.1%) granting conditional data accessibility, 12 provided informative answers regarding this topic (Question C3.1; see Figure 1). They generally referred to national and international legal frameworks (e.g. European legal framework, Material Transfer Agreement, the Health Information Privacy and Portability Act) and agreements. Interestingly, only one biobank highlight  role of the privacy guarantor for personal data protection in this procedure and only two biobanks (7.7%) provide the specific web link where it is possible to find documents regarding the legal framework (Question C3.2; see Figure 1).

Discussion

The culture of open science has began to spread over the past decade in different fields of life sciences (see Destro Bisol et al., 2014 and related citations therein). More specifically, sharing of scientific resources is increasingly perceived by scholars and researchers as a primary requirement for the development of new opportunities for collaboration (e.g. see Foster and Sharp, 2007; Fischer and Zigmond, 2010; Boulton, et al 2012; Mauthner and Parry, 2013). In the case of research involving human subjects, data and sample sharing practices have been carried out following different protocols, all of them facing obstacles and restrictions due to both practical (e.g. setting of informed consent) and ethical issues (e.g. privacy and confidentiality concerns, prediction of potential reuses) (see Blumenthal et al., 2006; Teeters et al., 2008). Moreover, given the different nature of data and samples, they does not necessarily follow identical sharing procedures. In fact, while data sharing culture in biosciences seems to be catching on among both researchers and policymakers, the same cannot be said for samples (Pereira, 2013). This is particularly true in the case of research biobanks where the finite nature of samples complicates their free circulation. Moreover, biobanks face the issue of operating (and often cooperating) in different countries where privacy laws not always coincide (Dove, Knoppers & Zawati, 2013). Starting from these premises, we conducted a questionnaire-based survey in order to shed light on how and at what level data and biological samples stored in research biobanks are accessible and reusable.

Primarily, most of the biobanks who responded to our survey give access their samples and data. However, a free and unconditioned accessibility is not quite a common practice. In fact, external research groups eager to use biobanks' resources must satisfy specific conditions in order to receive samples and get access to databases. However, most of the contacted biobanks provided

vague or not easily readable information about their accessibility criteria. This represent itself a non trivial result which shows that there is still little clarity, if not reluctance, in making sharing of scientific resources easier. This hindering in sharing is in sharp contrast with the emerging dependence of biomedical research on the activities of biobanks and the new ways of collaboration among researchers within global networks and consortia (Kaye, 2011; Kaye et al., 2015) as well as, in contradiction with the latest European research programme, Horizon 2020, where specific policies for open data and open access are envisaged (Leonelli, Spichtinger & Prainsack, 2015). However, the analysis of the informative answers points to three major issues related to the accessibility of biobanks resources.

Firstly, applicants are requested to explain what they would like to do with the required resources. Knowing this information is closely related with the specific data/bio-specimens sharing clause reported in the original consent form. At the same time, it provides a certain degree of control by the biobanks over the credentials and scientific reputation of the user as well as of his research group. Verifying reliability and seriousness of applicants and minimizing the misuse of data and samples is a fundamental requirement for an effective organizational impact of biobanks. It is both an ethical and technical approach of scientific resources' management which can promote public trust in the work of these institutions, thus increasing the willingness to participate to their activities (De Robbio, 2010).

Secondly, the role research funds, public or pri, in the establishment of fruitful collaborations between biobanks and research labs. Public research seems be preferred over studies granted by private funding bodies. This result is in compliance with the recommendation towards Open Access of scientific resources produced with public funds proposed by the Organisation for Economic Co-operation and Development (OECD) in his report "Promoting Access to Public Research Data for Scientific, Economic and Social Development" published in 2007. In the

OECD's report emerges the social, non epistemic, value of "public good" of sharing and the consideration of public scientific research as and investment. However, the concept of sharing as "public good" has been explicitly used by just one of the surveyed biobanks. On the other hand, sharing of scientific resources has also epistemic values regarding scientific rigor in favor of scientific progress, fostering an "effective biobanks knowledge generation" (Demir and Murtagh, 2013). Not only the source but also the availability of funds to carry out the research and to pay for the access to samples and data, results to be a fundamental criterion adopted by biobanks in order to disclose their resources to third parties. In fact, the presence of clauses directly related to economic benefits for biobanks discloses their possible "second nature" of for-profit institutions offering services of collection and storage of biological samples and, at the same time, giving the opportunity for researchers to access samples and data. Thus, we can assume a relationship between private funds, buying and selling of biobanks' resources and widespread sharing of data and biological samples. However, it is not clear if commercial nature of biobanks is really a barrier to sharing. Caufield et al (2014) suggest that sharing of data and samples is a practice that "may be impacted or hindered by the introduction of private funding and collaboration with private entities, as the expectations of private entities and agreements governing such partnerships may create sharing barriers". Differently, other authors hold that the increased mobility of data is unavoidably tied to their commodification and that venture capitalism, in its inflating and deflating of expectations, is interested in open data, overall when it does not lead to scientific insight which makes financial investments in this area both risky and potentially rewarding (Leonelli, 2013).

Thirdly, we found that recognition of co-authorship is a requirement for some biobanks in order to give access to their data. Co-authorship on publication as fair condition for the use of data produced by others has been also reported by Tenopir et al (2011) in their study on the data

sharing practices and perceptions by scientists. A similar result has been also found by Milanovic, Pontille and Cambon-Thomsen (2007) in their empirical study on sharing of biological samples and data in biosciences. This kind of request falls within the broader context of the management of scientific resources in order to gain advantages in the academic competition. According to Vogeli et al (2006), this behavior may contribute to spread a climate of mistrust and lack of cooperation within the scientific community.

To sum up, these results suggest that economic and academic aspects are involved in determining the extent of sharing of samples and data stored in biobanks. There is an *adage* so that if biobank's professionals mostly think to commercial pursuits, researchers mostly think to academic pursuits (Pereira, 2013). Fortunately, these detrimental attitudes for the scientific progress and for the ethics of science cannot be generalized. Recent empirical studies on data sharing have so far revealed the existence of at least one case of good practice. Anagnostou et al (2015) analyzing the data sharing rates in human paleogenetics showed that among researchers of this research field, data sharing is indeed a common practice. In fact, almost the totality of data were actually immediately available (97.6% of datasets). According to the authors it seems that this good sharing practice is to be attributed much more to a general awareness of the importance of openness and transparency for scientific progress rather than comply with norms or expectations of any scientific reward. Furthermore, Pereira (2013) depicts a more optimistic view about the willingness to share biological resources by biobanking professionals, highlighting that they "showed considerable interest in advancing research and a generally altruistic perspective toward sharing samples and making materials accessible to the research community".

One good practice could consist in disclosing the funding's origin and research's aims, in an era in which characteristics of public research are more and more similar to those of private commercial science. It is useful to keep in mind that the 'bank' metaphor overcome the notion of

"bio-repositories" or "bio-libraries" (Schneider, 2008) and that biobanks can diverge diametrically in scopes and outcomes, or diametrically divergent visions and practices can coexist within the same biobank. In this respect, standards to be applied to biobanks often fall under national and communitarian laws regulating trade, where often real and proper commercial battles are present and where bilateral agreements taken on global scale could influence or divert local interests and economies (e.g. the application of the Transatlantic Trade and Investment Partnership within national health services).

As a second step of this study, we investigated the reasons for the observed high heterogeneity of the requirements for the access to the biobanks' resources. Most of the surveyed biobanks adopted specific legal frameworks that researchers should take into consideration in order to gain access to samples and data. Comparing the information obtained from the biobanks, neither strategies nor standards result to be shared among these institutions. The different modalities of resource accessibility seem to be highly influenced by social context and legislations of the countries where biobanks operate. The fact that only few biobanks provided informative answers about this topic could be interpretable as strong evidence that resource sharing is still a cumbersome practice. This lack of clarity raises both ethical and practical issues: how to implement the sharing of ethical conditions linked to exploitation of data and biological samples? A first practical step could be represented by the opportunity for donors to make their own choices through the informed consent process. The ethical principles at the basis of informed consent in research involving human subjects (i.e. respect for persons, individual autonomy, protection of privacy) are inalienable and their importance is even more evident in the case of biobanks due to their nature of institutions involving multiple researchers within multiple research projects (Fullerton and Lee, 2011). But, precisely due to their nature, "it is difficult to obtain consent for all future research uses at the time of recruitment into the biobank or before

such research commences, as required in the original formulations of the Declaration of Helsinki" (Kaye et al., 2015). As stated by Jane Kaye et al (2015), classical informed consent fails to be an efficient tool to overcome the obstacles in data and samples sharing due to its static, paper-based format mostly recognized at national level. Furthermore, we must bear in mind that, particularly in the European context, privacy laws make the possibility to reuse data and samples extremely difficult . Interesting proposals coming from the Anglo-Saxon world regard the possibility for participants to establish, through Information and Communication Technologies (ICT), an ongoing, bidirectional communication with biobank institutions to refresh or withdraw their consent for new research projects (Stein and Terry, 2013; Kaye et al., 2015). In such a dynamic consent, authorization of individuals on handling of their personal data could travel with the same datasets containing biological and personal data (Terry et al., 2013). The dynamic consent approach could be conceived as a manner to preserve the individual right to decide in autonomy on the basis of the information received (participants could be informed on projects aims and methods as much as they want) and, at the same time, as a manner to protect individuals' privacy (each participant is free to handle and authorize flows of personal data and to know regulations on data protection). Nevertheless, dynamic consent approach and strict laws on data protection are not useful in all cases. Indeed, privacy laws and the rising attention on individualistic rights can hinder the broad informed consent model and, overall, can hinder those bio-repositories usually established to protect collective rights such as public health. For instance, cancer registries or retrospective studies could be damaged by the strict rules proposed by the European Parliament's resolution (12 March 2014) on privacy which holds the need of asking participants their consent for every new research project involving their data and samples (Casali, 2014).

This reasoning leads back to the aforementioned problem of the lack of common and standardized operating procedures (SOPs) and heterogeneity in access rules. In fact, this

fragmentation not only limits the benefits for the academia, but it can also be seen as one of the reasons why for prospective donors is difficult to understand the role and the activities of biobanks. Undoubtedly, advancements of biomedical research are strictly linked with the increase of public interest towards biobanks' activities: trivially, without donors they cannot operate. But the willingness to donate its own sample and actively participate to biobanks activities are in turn strictly linked with the clarity in exposing the importance to participate in medical research (i.e. benefits deriving from biobank research) and the manner in which biological samples and data will be used and made available to the scientific community. In short, the only way to undertake mutually productive relationships between donors and biobanks is through trust "understood as something which demands knowledge and consent" (Richter, 2012).

Concluding remarks

In this paper, we have attempted to analyze the degree of accessibility and reusability of data and biological samples stored in research biobanks following an empirical approach. Mainly, this study suggests that, in spite of general consensus of scientific community on the importance of open access of scientific resources, sample and data sharing barriers are still acting among biobanks and researchers. Undoubtedly, this preliminary investigation need to be continued and improved in order to support (or even call into question) the results obtained. Particularly, increasing the number of surveyed biobanks and the related differences of socio-cultural contexts could help in producing a more detailed picture of sharing behaviors and their differences on the basis of countries where biobanks operate. Furthermore, a greater extent of information could be obtained following a two-step research protocol based on quantitative approach as those used in the present study, and a second, more deeply focused, qualitative investigation (e.g. semi-structured interviews, focus groups and interviews) on the main issues emerged from the first

step. According to Mertz et al (2014), empirical approaches provide an opportunity to overcome the classical descriptive aim of social science methods applied in studying the scientific environment. From this point of view, the so-called “empirical ethics” (see Hope, 1999 and Molewijk et al., 2004) may contribute to increase the knowledge on how and in what way all the agents involved in the life cycle of biomedical research share their work, encouraging the full exploitation of their scientific products.

Acknowledgements

We would like to thank the biobanks and their anonymous members who contributed to the survey. This work was supported by the Istituto Italiano di Antropologia (http://www.isita-org.com/).

Authors’ contributions

Conceived and designed the research: MC, PA, GM and FR; collected the data: MC, PA and GM; analysed the data: FDA, MC, PA and VD; wrote the paper: MC, FDA, PA and FR, with critical and theoretical inputs from GDB. All authors read and approved the final manuscript.

References

Anagnostou P, Capocasa M, Milia N, Sanna E, Battaggia C, Luzi D, Destro Bisol G. 2015. When data sharing gets close to 100%: what human paleogenetics can teach the Open Science movement. *PLoS One* 0:e0121409. DOI: 10.1371/journal.pone.0121409.

Arzberger P, Schroeder P, Beaulieu A, Bowker G, Casey K, Laaksonen L, Moorman D, Uhlir P, Wouters P. 2004. Promoting access to public research data for scientific, economic and social development. *Data Science Journal* 3:135-152.

Blee KM, Currier A. 2001. Ethics beyond the IRB: An introductory essay. *Qualitative Sociology* 34:401-413. DOI: 10.1007/s11133-011-9195-z.

Blumenthal D, Campbell EG, Gokhale M, Yucel R, Clarridge B, Hilgartner S, Holtzman NA. 2006. Data withholding in genetics and the other life sciences: prevalences and predictors. *Academic Medicine* 81:137–145.

Bohannon J. 2013. Genealogy databases enable naming of anonymous DNA donors. *Science* 339:262. DOI: 10.1126/science.339.6117.262.

Boulton G, Campbell P, Collins B, Elias P, Hall W, Laurie G, O'Neill BA, Rawlins M, Thornton DJ, Vallance P, Walport M. 2012. *Science as an open enterprise*. London: The Royal Society; 2012.

Cadigan RJ, Juengst E, Davis A, Henderson G. 2014. Underutilization of specimens in biobanks: an ethical as well as a practical concern?. *Genetics in Medicine* 16:738-740. DOI: 10.1038/gim.2014.38.

Casali PG. 2014. Risks of the new EU data protection regulation: an ESMO position paper endorsed by the European oncology community. *Annals of Oncology* 25:1458-1461. DOI: 10.1093/annonc/mdu218.

Caulfield T, Burningham S, Joly Y, Master Z, Shabani M, Borry P, Becker A, Burgess M, Calder K, Critchley C, Edwards K, Fullerton SM, Gottweis H, Hyde-Lay R, Illes J, Isasi R, Kato K, Kaye J, Knoppers B, Lynch J, McGuire A, Meslin E, Nicol D, O'Doherty K, Ogbogu U, Otlowski M, Pullman D, Ries N, Scott C, Sears M, Wallace H, Zawati MH. 2014. A review of the key issues associated with the commercialization of biobanks. *Journal of Law and the Biosciences* 1:94-110. DOI: 10.1093/jlb/lst004.

Colledge F, Persson K, Elger B, Shaw D. 2014. Sample and data sharing barriers in biobanking: consent, committees, and compromises. *Annals of Diagnostic Pathology* 18:78-81. DOI: 10.1016/j.anndiagpath.2013.12.002.

D'Abramo F. 2015. Biobank research, informed consent and society. Towards a new alliance?. *Journal of Epidemiology and Community Health* [Epub ahead of print]. DOI: 10.1136/jech-2014-205215.

De Robbio A. 2010. Biobanche e proprietà intellettuale: commons o caveau?. *Bibliotime* 14:3.

Demir I, Murtagh MJ. 2013. Data sharing across biobanks: epistemic values, data mutability and data incommensurability. *New Genetics and Society* 32:350-365. DOI: 10.1080/14636778.2013.846582.

Destro Bisol G, Anagnostou P, Capocasa M, Bencivelli S, Cerroni A, Contreras J, Enke N, Fantini B, Greco P, Heeney C, Luzi D, Manghi P, Mascalzoni D, Molloy J, Parenti F, Wicherts J, Boulton G. 2014. Perspectives on open science and scientific data sharing: an interdisciplinary workshop. *Journal of Anthropological Sciences* 92:179-200. DOI: 10.4436/JASS.92006.

Dove ES, Knoppers BM, Zawati MH. 2013. An ethics safe harbor for international genomics research?. *Genome Medicine* 5:99. DOI: 10.1186/gm503.

Fischer BA, Zigmond MJ. 2010. The essential nature of sharing in science. *Science and Engineering Ethics* 16:783-799. DOI: 10.1007/s11948-010-9239-x.

Foster MW, Sharp RR. 2007. Share and share alike: deciding how to distribute the scientific and social benefits of genomic data. *Nature Reviews Genetics* 8:633-639.

Fullerton SM, Lee SS. 2011. Secondary uses and the governance of de-identified data: lessons from the human genome diversity panel. *BMC Medical Ethics* 12:16. DOI: 10.1186/1472-6939-12-16.

Godard B, Ozdemir V, Fortin M, Égalité N. 2010. Ethnocultural community leaders' views and perceptions on biobanks and population specific genomic research: a qualitative research study. *Public Understanding of Science* 19:469-485.

Haga SB, Beskow LM. 2008. Ethical, legal, and social implications of biobanks for genetics research. *Advances in Genetics* 60:505-544. DOI: 10.1016/S0065-2660(07)00418-X.

Henderson GE, Cadigan RJ, Edwards TP, Conlon I, Nelson AG, Evans JP, Davis AM, Zimmer C, Weiner BJ. 2013. Characterizing biobank organizations in the U.S.: results from a national survey. *Genome Medicine* 5:3. DOI: 10.1186/gm407.

Hirtzlin I, Dubreuil C, Preaubert N, Duchier J, Jansen B, Simon J, Lobato De Faria P, Perez-Lezaun A, Visser B, Williams GD, Cambon-Thomsen A; EUROGENBANK Consortium. 2003. An empirical survey on biobanking of human genetic material and data in six EU countries. *European Journal of Human Genetics* 11:475-488.

Hope T. 1999. Empirical medical ethics. *Journal of Medical Ethics* 25:219.

Kaufman DJ, Murphy-Bollinger J, Scott J, Hudson KL. 2009. Public opinion about the importance of privacy in biobank research. *American Journal of Human Genetics* 85:643-654. DOI: 10.1016/j.ajhg.2009.

Kaye J. 2006. Do we need a uniform regulatory system for biobanks across Europe?. *European Journal of Human Genetics* 14:245-248.

Kaye J. 2011. From single biobanks to international networks: developing e-governance. *Human Genetics* 130:377-382. DOI: 10.1007/s00439-011-1063-0.

Kaye J. 2012. The tension between data sharing and the protection of privacy in genomics research. *Annual Review of Genomics and Human Genetics* 13:415-431. DOI: 10.1146/annurev-genom-082410-101454.

Kaye J, Whitley EA, Lund D, Morrison M, Teare H, Melham K. 2015. Dynamic consent: a patient interface for twenty-first century research networks. *European Journal of Human Genetics* 23:141-146. DOI: 10.1038/ejhg.2014.71.

Knoppers BM. 2010. Consent to 'personal' genomics and privacy. *EMBO Reports* 11:416-419. DOI: 10.1038/embor.2010.69.

Lemke AA, Wolf WA, Hebert-Beirne J, Smith ME. 2010. Public and biobank participant attitudes toward genetic research participation and data sharing. *Public Health Genomics* 13:368-377. DOI: 10.1159/000276767.

Leonelli S. 2013. Why the current insistence on Open Access to scientific data? Big data, knowledge production, and the political economy of contemporary biology. *Bulletin of Science, Technology & Society* 33:6-11. DOI: 10.1177/0270467613496768.

Leonelli S, Spichtinger D, Prainsack B. 2015. Sticks and carrots: encouraging open science at its source. *Geography and Environment* [Epub ahead of print]. DOI: 10.1002/geo2.2.

Levine C. 1988. Has AIDS changed the ethics of human subjects research?. *Journal of Law, Medicine & Ethics* 16:167-173. DOI: 10.1111/j.1748-720X.1988.tb01942.x.

Macilotti M. 2010. Le biobanche: disciplina e diritti della persona. In: Rodotà S, Tallacchini M, eds. *Trattato di biodiritto*. Milano: Giuffrè, 1195-1215.

Macilotti M. 2013. Informed consent and research biobanks: a challenge in three dimensions. In: Pascuzzi G, Izzo U, Macilotti M, eds. *Comparative issues in the governance of research biobanks*. Berlin Heidelberg: Springer, 143-161. DOI: 10.1007/978-3-642-33116-9_9.

Mauthner NS, Parry O. 2013. Open Access digital data sharing: principles, policies and practices. *Social Epistemology* 27:47-67. DOI: 10.1080/02691728.2012.760663.

Mertz M, Inthorn J, Renz G, Rothenberger LG, Salloch S, Schildmann J, Wöhlke S, Schicktanz S. 2014. Research across the disciplines: a road map for quality criteria in empirical ethics research. *BMC Medical Ethics* 15:17. DOI: 10.1186/1472-6939-15-17.

Milanovic F, Pontille D, Cambon-Thomsen A. 2007. Biobanking and data sharing: a plurality of exchange regimes. *Genomics, Society & Policy* 3:17-30.

Molewijk B, Stiggelbout AM, Otten W, Dupuis HM, Kievit J. 2004. Scientific contribution. Empirical data and moral theory. A plea for integrated empirical ethics. *Medicine, Health Care and Philosophy* 7:55-69. DOI: 10.1023/B:MHEP.0000021848.75590.b0.

Nelson B. 2009. Empty archives. *Nature* 461:160-163. DOI: 10.1038/461160a.

Pereira S. 2013. Motivations and barriers to sharing biological samples: a case study. *Journal of Personalized Medicine* 3:102-110. DOI: 10.3390/jpm3020102.

Richter C. 2012. Biobanking. Trust as basis for responsibility. In: Dabrock P, Taupitz J, Ried J, eds. *Trust in biobanking. Dealing with ethical, legal and social issues in an emerging field of biotechnology*. Berlin-Heidelberg: Springer, 43-66.

Rodriguez LL, Brooks LD, Greenberg JH, Green ED. 2013. The complexities of genomic identifiability. *Science* 339: 275-276. DOI: 10.1126/science.1234593.

Schneider I. 2008. 'This is not a national biobank...': the politics of local biobanks in Germany. In: Gottweis H, Petersen A, eds. *Biobanks: governance in comparative perspective*. Abingdon: Routledge, 88-108.

Stein DT, Terry SF. 2013. Reforming biobank consent policy: a necessary move away from broad consent toward dynamic consent. *Genetic Testing and Molecular Biomarkers* 17:855-856. DOI: 10.1089/gtmb.2013.1550.

Teeters JL, Harris KD, Millman KJ, Olshausen BA, Sommer FT. 2008. Data sharing for computational neuroscience. *Neuroinformatics* 6:47-55. DOI: 10.1007/s12021-008-9009-y.

Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, Manoff M, Frame M. 2011. Data sharing by scientists: practices and perceptions. *PLoS One* 6:e21101. DOI: 10.1371/journal.pone.0021101.

Terry SF, Shelton R, Biggers G, Baker D, Edwards K. 2013. The haystack is made of needles. Genetic Testing and Molecular Biomarkers 17:175-177. DOI: 10.1089/gtmb.2012.1542.

Vogeli C, Yucel R, Bendavid E, Jones LM, Anderson MS, Louis KS, Campbell EG. 2006. Data withholding and the next generation of scientists: results of a national survey. *Academic Medicine* 81:128–136.

Weijer C, Goldsand G, Emanuel EJ. 1999. Protecting communities in research: current guidelines and limits of extrapolation. Nature Genetics 23:275-280.

Wolf SM, Crock BN, Van Ness B, Lawrenz F, Kahn JP, Beskow LM, Cho MK, Christman MF, Green RC, Hall R, Illes J, Keane M, Knoppers BM, Koenig BA, Kohane IS, Leroy B, Maschke KJ, McGeveran W, Ossorio P, Parker LS, Petersen GM, Richardson HS, Scott JA, Terry SF, Wilfond BS, Wolf WA. 2012. Managing incidental findings and research results in genomic research involving biobanks and archived data sets. Genetics in Medicine 14:361-384. DOI: 10.1038/gim.2012.23.

1 Zika E, Paci D, Braun A, Rijkers-Defrasne S, Deschenes M, Fortier I, Laage-Hellman J, Scerri
2 CA, Ibarreta D. 2011. A European survey on biobanks: trends and issues. *Public Health*
3 *Genomics* 14:96-103. DOI: 10.1159/000296278.

4

5

Table 1. Geographic distribution of biobanks involved in this study. Percentage of respondents in brackets.

Continent	Country	Biobanks Invited	Biobanks Responded
Africa	Sud Africa	1	0 (0)
	Zimbabwe	1	0 (0)
America	Brazil	1	0 (0)
	Canada	12	2 (16.7)
	USA	91	14 (15.4)
Asia	China	2	0 (0)
	India	3	0 (0)
	Iran	1	0 (0)
	Israel	3	1 (33.3)
	Japan	4	1 (25.0)
	Korea	1	0 (0)
	Malaysia	3	0 (0)
	Qatar	1	0 (0)
	Singapore	4	0 (0)
	Taiwan	1	0 (0)
	Thailand	2	0 (0)
Europe	Austria	4	2 (50.0)
	Belgium	3	1 (33.3)
	Estonia	1	1 (100)
	Finland	1	1 (100)
	France	7	2 (28.6)
	Germany	19	3 (15.8)
	Greece	2	0 (0)
	Hungary	1	0 (0)
	Iceland	1	0 (0)
	Ireland	3	1 (33.3)
	Italy	12	5 (41.7)
	Latvia	1	0 (0)
	Luxembourg	1	0 (0)
	Malta	1	0 (0)
	Netherlands	4	1 (25.0)
	Norway	2	1 (50.0)
	Poland	1	0 (0)
	Portugal	1	0 (0)
	Spain	5	1 (20.0)
	Sweden	5	1 (20.0)
	Switzerland	4	0 (0)
	United Kingdom	16	6 (37.5)
Oceania	Australia	12	2 (16.7)
	Total	238	46 (19.3)

Figure 1. Informativeness of the answers given to the open questions.