Traceability as a means in data protection of research Biobanks

A. Barmpagalou[a,b], M. Manousos[c], A. Perdikaris

(a) National and Kapodistrian University of Athens, School of Law, Economics and Political Sciences, Faculty of Law, Athens, Greece

(b) European Institute for Evolution and Integration, Messinia, Greece

(c) Foodstandard S.A., Athens, Greece

Abstract

IT is a critical tool for data storage, acquisition and analysis; however the approach in different biobanks varies. Pseudoanonimization on the other hand is a means for exchanging a range of personal identifier for an indirect identifier, which is usually a code key. However even a code key is not sufficient to dissociate the data and the sample, consequently the donor, due to the fact that other indirect information, such as place, collection time or DNA, could make the sample identical. Thus a coding system is required and different levels of security related to code keys should be in place. In the proposed system the type of donor's consent is taken into account in order to define the different access levels to the data base. A special provision should also be taken in case the donor wishes not to be informed of incidental health findings. The more specific is the given consent the less autonomy is provided regarding the possible sample uses. The architecture of the system is based on a traceability system, where permissions to operators wishing to follow back the sample origin, is given according to their role in the sample treatment and the donor's given consent.

Key words: Biobanks, Genetic, Data, Security, IT, Donor, anonymisation, Filemaker, WebDirect

Introduction

The legal definition of biobank is found in Council's Recommendation (2006)4 on research on biological materials of human origin, more specifically in article 17, where it is stated that a population biobank is:

"A collection of biological materials that has the following characteristics:
The collection has a population basis;

I. It is established, or has been converted, to supply biological materials or data derived therefrom for multiple future research projects;
II. It contains biological materials and associated personal data, which may include or be linked to genealogical, medical and lifestyle data and which may be regularly updated;
III. It receives and supplies materials in an organised manner."

Biobanks identified by the collection of biological samples and the data associated with them. Both elements are organized in a systematic way for use by authorized users associated with different access levels. As a concept biobanks are under continuous development due to the respective evolution of medical science, sample preservation and IT technology achievements. Additionally new uses and sample treatments have been invented and new biobanks continuously established [European Commission, 2012]. As sample and data "collectors", biobanks collect and store data on a regular basis and perform research activities continuously, thus the collected samples might be used several times. In order to secure the anonymisation of the donors a series of processes is applied [European Commission, 2012], [Beier et al., (eds.), 2011].

IT technology developments have given an extraordinary boost to biobanking growth and provided a more structured scientific work frame establishing in parallel a research network on a global base [Kaye et al., 2009]. Current research in medical research, involves scientists, experts and staff from several scientific, expertise and technical backgrounds. These teams and bodies (labs, hospitals, universities, etc.) should perform a research activity under the same network, which is not an easy task, if there is not a central coordinating system. The same time this system should be able to secure the stored data and comply with the requirements of the data protection legislation.

Biobank structure

The number of samples stored and manipulated in a biobank is the core element that identifies its type and scope [Asslaber et al., 2007], [Riegman et al., 2008]. Usually smaller biobanks are directed to research purposes only. According to the European Commission's report on Biobanks for Europe, biobanks could be classified into six (6) different formats, depending on the samples they store and the intended use [European Commission, 2012]:

- **Population**-based biobanks
- **Disease-oriented** biobanks
- **Case-control** biobanks
- **Tissue** banks
- Biobanking within the context of **clinical trials**
- Other **specific** biobanking formats: Guthrie cards,cord blood, stem cells

The first three basically are oriented to disease research, where the first kind usually store samples from healthy donors, the second stores biological material and addressed to clinical care and the third is referring to samples coming from donors with a specific disease [European Commission, 2012]. The last three are determined by their title.

Anonymisation or Pseudonymisation

Anonymisation or Pseudonymisation can be achieved by substituting identifiers related to a donor with an alphanumerical code for example or a code key. This means offer a protection of privacy, but there is always a risk for future accidental or not identification. One of the advantages of this process is that even if the donor withdraws, the sample, taking into account that is anonymized, might not be destroyed by the administrator.

The above process should protect the donor and the same time ensures that a traceability system is in place. Traceability is very important in biobanking for two basic reasons:

1. The first is related to the research itself and usability of the sample and
2. The right of the donor to demand his sample been destroyed or delete any related identification marker.

On the other hand a process should be established in order to balance donor's requests and restrictions, information to donor on incidental health findings and donor's wish not to be informed in such case.

Obviously the system as described is too complicated to be regulated without a specific IT system to control it, where all the parameters will be set from the beginning of sample storage and the related parameters set up in the biobank database. The system should also take into account the existing legislation, which is not always so extended to cover such issues, especially when the biobanking network is global.

Consent

Elements of human body can be processed only if there is an informed consent, which is a core requirement in all kinds of medical research. There are different definitions of consent depending on the existing legislation in different countries.

Consent may include except sample use, property rights, sample processing and genetic data processing at a second stage.

If we examine the issue from an IT point of view and follow the approach that consent procedures is a continuum from blanket consent to consent to a specific body, then we end with a frame work where Blanket consent means more autonomy and in contrary specific consent means more specificity and less autonomy [Hansson, M.G., et al., 2006]. This approach and a traceability system was a core element in order to develop the IT product for biobank management, described in the last section.

IT and bioinformatics

As mentioned IT tools are core elements in a biobanking system allowing the management, processing and analysis of data. A traceability system cannot be established and data acquisition cannot take place without and IT management system. IT also regulates the sample collection and storage system and the information associated with them, produced during the recording process as well as during the scientific research.

In recent research by the European Commission, it was stated by several biobank managers that, one of the main obstacles in research and share of sample tissues is the absence of applications for databases with security rules, function under a framework defined by law and IT science [European Commission, JRC, 2010]. With an extended IT system it will be possible to connect different biobanks and exploit a wider range of findings and different donors' histories and associate their data in a more comparable way.

IT application

The application that we propose is called "itisseu" (www.itiss.eu) and developed by using the "Filemaker" platform, which characterised by its breakthrough web

technology for data access anywhere. With the “Filemaker” platform we will be able to combine the power of a desktop application with the simplicity of a web browser to manage our information. “Filemaker” WebDirect is breakthrough web technology that runs custom business solutions directly in a web browser on a desktop or laptop - with no special web development skills required. With “Filemaker” WebDirect, we will not need to use coding tools such as PHP, HTML5, CSS or JavaScript to create the application for the web.

Since “Filemaker” WebDirect looks like and acts like a desktop application, it will give us familiar functionality for the construction of our application such as desktop-style interaction for use themes, styles, charts, menus and many more. It will even give us the opportunity to drag and drop files into container fields or get live updates which means that we will be able to get instant access to changes in our data with no need to refresh the browser.

Through its automated processes we will be able to enable scripts, calculations and conditional formatting to validate data and streamline workflow. The application that will be created with this platform will be accessed by the use of tablet and windows and mac operation systems.

The “Filemaker” platform is ideal for the construction of this application, as it will assure our data with the latest industry- standard security. “Filemaker” will secure our data wherever we store it, e.g. desktop, device or server. Using powerful AES 256-bit encryption, our data will be secure from unauthorized access when stored in a “Filemaker” client or hosted on “Filemaker” Server and for that reason “Filemaker” Pro 13 Advanced is required to enable encryption. The platform will also support audit log that will be recording all the information of the users logging in and the changes that they make so it will function as a full traceability system for the application and the database.

Our team has established an internet based database where the administrator will upload all the data of the donor, which will be encoded accordingly to the donor’s consent. All the users of the database will be divided into groups according their accessibility depending on their authorization level. Each group will also be able to

upload new data (i.e test results, issues of health etc) which will also be visible to other users within the constraints set to each user group. The data that will be at every ones disposal are the main characteristics of the donor such as the gender, age group and object of study. Each sample will also be accompanied with a unique identification Number Database (IND), originated in a way that it will be impossible the samples to be mistaken. This unique identification number both will ensure the full traceability of the samples and eliminate the unauthorized access to the data.

Conclusion

The described application (itisseu) is a friendly user application and fully customized, having the flexibility to integrate any legal or IT demand. Consequently some of its basic advantages are:

1. It is a highly customizable web application.
2. Easily amended and updated in order to support in the future new sample types and integrate new legislative rules.
3. Ability for biobanks working in a common network to add custom processes without having to develop a new application
4. “Filemaker” as a platform is very popular with the appropriate support.
5. Could promote biobank networking under a common management tool.
6. Full traceability and recording support to capture all amendments and updates to sample data.

The tool can be used for any kind of research biobank and is willing to combine the legislative demands, the donors’ rights and the effective management of samples and data associated with them, under a secure frame.

References

Asslaber, M., Zatloukal, K., (2007), Biobanks: transnational, European and global networks, Briefings in Functional Genomics and Proteomics, Vol. 6, No. 3, p. 193-201.

Beier, K., Schnorrer, S, Hoppe N. (eds.), (2011), The Ethical and Legal Regulation of Human Tissue and Biobank Research in Europe; Proceedings of the Tiss.EU Project. University of Gottingen, ISBN 978-3-86395-031-6

European Commission, (2012), Biobanks for Europe; A challenge for governance, EU Publications Office, ISBN 978-92-79-22858-2

European Commission, JRC, (2010), Biobanks in Europe: Prospects for Harmonisation and Networking, EU Publications Office, ISBN 978-92-79-15783-7

Hansson, M.G., et al., (2006), Should donors be allowed to give broad consent to futurebiobank research? Lancet Oncol., 7(3): p. 266-269

Kaye, J., Heeney, C., Hawkins, N., de Vries, J., Boddington, P., (2009), Data sharing in genomics – re-shapingscientific practice´, Nat. Rev. Gene, Vol. 10, pp. 331-335

Riegman, P.H., Morente, M.M., Betsou, F., de Blasio, P., Geary, P., (2008). Biobanking for better healthcare andthe Marble Arch International Working Group on Biobanking for Biomedical Research, MolecularOncology, Vol. 2, No. 3, pp. 213-222.