Adapting research to the 21st century – the Swedish Biobank Register

L. Norlin[1], M. Fransson[1], S. Eaker[2], G. Elinder[3] and J.-E. Litton[1]

1) Karolinska Institutet, Department of Medical Epidemiology and Biostatistics, Stockholm, Sweden
2) Regionalt biobankscentrum in Uppsala/Örebro health care region, Uppsala University Hospital Uppsala, Sweden
3) Karolinska Institutet, Department of Clinical Science and Education, Södersjukhuset, Stockholm, Sweden

Correspondence: Loreana Norlin, Karolinska Institutet, Department of Medical Epidemiology and Biostatistics, Box 281, SE-171 77 Stockholm, Sweden
E-mail: loreana.norlin@ki.se Telephone: +46 8 524 87408

ABSTRACT

In Sweden, there are currently nearly 600 biobanks. The Swedish Biobank Register (SBR) is an on-going national investment by the county councils working to capture information in one database about all biobank samples collected from patients attending the Swedish medical health care. The aim of the SBR is to gather enough information about biobank samples to be able to physically trace the samples.

The BioBanking and Molecular Resource Infrastructure of Sweden (BBMRI.se) has been given the task of extending the SBR Information System (IS) with functionality useful for research in connection to health care, quality registers and large patient cohorts. The research extension is called BBMRI catalogue over sample collections for research. To achieve this, the SBR-IS will be extended with attributes useful for both epidemiological and clinical research enabling authorized researchers to search for samples stored at non-clinical biobanks nationwide. The Swedish Biobank Register, together with the BBMRI research catalogue, will be a unique resource for research. SBR-IS will contain information about biobank samples collected by both clinical and population-based biobanks specifically established for research purposes but BBMRI.se researchers will only be granted access to data related to population-based biobanks. As BBMRI.se is the Swedish hub of the pan-European biobank project BBMRI, whose goal is to promote excellence and efficacy in European life science research, the BBMRI research catalogue will also be made compatible with the European register by applying its minimum data set describing biobanks and their objects. In this paper we describe the implementation. Our belief is that it will pave the way for connecting biobanks on an international level as well as stimulate collaborations and optimize usage of biobank samples. In the long run, patients and sample donors will benefit as new results with high statistical power emerge from large scale studies.

BACKGROUND

The trend within biomedical research is shifting towards large scale studies. It is often discussed that biobanking poses a new strategic cornerstone in research and health care [1]. However, with a small-scale biobank structure it is difficult to participate in large scale national and international collaborative studies, one of the reasons being the lack of harmonization of sample collection and data handling. Research conducted by the National Cancer Institute (NCI) in the U.S. shows that finding and obtaining high-quality samples from cancer patients remains a major obstacle in performing translational research [2]. Moreover, only large scale studies have the possibility to ensure adequate statistical power for data analysis. Large-scale biobanks might be the only time- and cost-effective option to provide researchers with biospecimens collected using standardized methods with which to conduct large scale research [3]. So far, most biobank infrastructure initiatives at the European level have developed as organisational networking, pure harmonization efforts in terms of specific data sets or standardising operational procedures for material collection, fixation, and storage. Whilst it is important to raise awareness of the diversity in the field, the full potential of biobank infrastructures will only be realised if national biobank registries are constructed. Researchers, sample donors and research itself would benefit considerably if a unifying IT platform existed that harmonizes data from different biobanks. As a first step, harmonization will cover information about biobank organisations using the same format for describing the biobanks. As a next step, aggregated data relating to studies can be harmonized, and finally the objects of a biobanks (donors and samples).

A NATIONAL REGISTER FOR BIOBANK SAMPLES

In Sweden, there are nearly 600 biobanks where biobank samples are managed [4]. The handling of biobank samples in Sweden currently conforms to the Swedish Biobanks in Medical Care Act (2002:297) [5]. The main purpose of the act is to protect donor integrity and states that a donor, after having consented for his or her sample to be stored in a biobank, has the possibility to later withdraw or restrict this informed

consent. The withdrawal or restriction is done by the biobank by either destructing the sample or anonymizing it, irreversibly breaking the link to the donor. To fulfill this, knowledge about where a biobank sample is physically stored at any given point in time is a prerequisite. In order to facilitate the administration of biobank samples the Swedish Association of Local Authorities and Regions (SALAR) has purchased a Laboratory Information Management System (LIMS). The LIMS, called the Swedish Biobank Register Information System (SBR-IS), is web-based and is used as a national register for all biobank samples collected within the public health care [6]. Sample data in the system is based on the unique personal identity number (PIN) of every Swedish citizen [7]. Only data that is necessary to identify the samples belonging to specific donors is stored in the Information System. A new act that is under development might result in comprising all biobank samples, not only the ones taken within the health care sector.

IMPROVING THE USAGE OF BIOBANK SAMPLES IN RESEARCH

The BioBanking and Molecular Resource Infrastructure of Sweden (BBMRI.se) has a work package partly devoted to extending the SBR-IS and making it useful for research [9]. In addition to serving the purpose of traceability using donor personal identity numbers (PIN), the Swedish Biobank Register will be extended to include a research catalogue (the BBMRI catalogue over sample collections for research) which will contain information about non-clinical biobanks that collect biospecimens and data for research purposes. The goal is to create a tool that promotes research collaboration between distinct research groups and to gain new knowledge by improving the usage of biobanks. BBMRI.se, the hub of the pan-European biobank network BBMRI, was given the task of investigating how the Swedish Biobank Register could potentially be used in research regarding health care, quality registers, large patient cohorts etc. without putting donor integrity at risk. This task was given to BBMRI.se by SALAR and the Swedish Research Council. The extended information system should support four use cases A-D:

A) Search for samples for donors that wish to change a previously given consent
B) Search for biobank samples of a certain criteria
C) Search for individuals of a certain criteria
D) Search for biobank samples from specific individuals

See section "Examples of use cases" for detailed examples from each use case.

Use cases A and D are already supported by the SBR-IS whereas functionality for use cases B and C needs to be added. In order to fulfill these use cases, the SBR Information System needs to be extended with additional attributes and search capabilities. The additional attributes presented in table 1 describe biobanks, samples collections/studies, donors and samples.

Authorized researchers who are part of the BBMRI.se infrastructure will have the ability to search and obtain data about biobank samples from BBMRI.se biobanks based on attributes described in table 1, e.g. to perform their own epidemiological or clinical studies or to find collaboration partners. Collecting the information from BBMRI.se researchers will pave the way for the following possibilities:

- stimulate research collaborations
- make existing sample collections visible to others
- enable statistical analysis on aggregated data for comparison between biobanks
- enable future assembly of object data

Examples of use cases

For all four use cases described, a BBMRI.se researcher would log into the system with his/her user name and password. Using a graphical user interface the researcher could search for and obtain information about the questions formulated in use cases A-D.

Example of Use Case A
"Find all biobank samples collected from a person under a given period of time that the person wishes to throw away."

In order to fulfill use case A, the system needs to hold information about who the donor is, (name and PIN), samples donated by that person (sample ID), when the samples were harvested (sampling date) and where the samples are (contact information for the biobank).

Example of Use Case B
"Search for sample collections with at least ten biopsies from thoracic aorta, where there are also blood, serum and plasma samples collected at the same time as the biopsy (± one week). See what clinical data have been registered for the donors such as blood pressure, lipids and BMI."

This use case implies that the system user primarily wants to find samples with a set of desired properties. In order to fulfill use case B, the system needs to contain attributes describing the samples such as where on the body the samples were taken (anatomical position) and what type of sample it is (material type). The system should also hold information about if the sample is part of a bigger sample collection (samples having the same properties) or a study (study ID) linking to information about what type of clinical data has been gathered about the donors. The search can be performed on the sample level or on the sample collection level.

Example of Use Case C
"Search for male donors that have prostate cancer or female donors that have breast cancer and where serum, plasma or tissue samples were donated before diagnosis. Find out how long before the diagnosis the

Table 1. An overview of the attributes needed to extend the Swedish Biobank Register Information System (SBR-IS) to support use cases for BBMRI.se and make the SBR-IS compatible with the BBMRI minimum data set [10].

Attribute	Data type	Comment
Data describing biobanks		
URL	Hyperlink	Complete http:// address
Country	String	ISO-standard 3166 alpha2, e.g. SE
Data describing sample collections and studies		
StudyName	String	Name of study in local language
StudyNameEnglish	String	Name of study in English
Contact	String	Contact person and address etc.
DescriptiveTitle	String	According to ethical application or such. No confidential information
StudyType	Single Select	E.g. case-cohort, case-control, cohort, longitudinal, other
TypeOfCollection	Single Select	E.g. population based, disease specific, other
DiseaseCode	String or multiple select	For disease specific studies, ICD-10 is proposed.
ClinicalData	Boolean	Multiple existence attributes revealing if information has been collected about the subjects. From the P3G.org research theme + BMI, Lipids, blood pressure
Planned_Sex	Integer	The gender of donors planned for the study
Planned_NumberofIndividuals	Integer	The number of donors planned for the study
Planned_AgeGroup	Integer	The age of the donors planned for the study, stated as a 5 year interval (compare to sample level)
Planned_SamplingPeriod	Date	Date of first and last sample harvested
Planned_MaterialType	String	The nature of the samples planned for the study (according to SPREC [9])
Planned_AnatomicalPosition	String	If material type is 'tissue'
Planned_OmicsData	String	E.g. genomics, proteomics, metabolomics etc.
Actual_NumberofIndividuals	Integer	Aggregated form based on the donor entity
Actual_NumSamples	Integer	Aggregated form based on sample level
CoolChain	Boolean	Indication if samples in the sample collection have been kept cold from sampling to storage
DOI	String	D.O.I. for publications using samples from this study, e.g. 10.1000/182.
Data describing donors		
None in addition to already existing attributes in SBR-IS		
Data describing samples		
AgeGroup	Single Select	Age of donor at sample collection stated as 5-year intervals e.g. 0-4 years, 5-9 years etc.
SPRECCode [9]	String	ISBER sample pre-analytical code
DiseaseCode	String or multiple select	ICD-10 is proposed. One sample can have up to five diagnoses
OmicsDataAvailable	String	Existence attributes revealing if information about -omics data have been gathered about the samples, e.g. genomics, proteomics, transcriptomics, metabolomics etc.

samples were collected. Find out if there are any fresh or frozen biopsy tissues for these donors."

This use case implies that the system user primarily wants to find individuals with a set of desired properties. This means searching on attributes that are directly linked to the donor instead of the samples stored in the biobank. Use case C is considered of lower relevance as it can easily be converted into a type B case and therefore has not been further developed. However, the system would then need to contain information about the diagnosis on the sample or the sample collection level.

Example of Use Case D

"Using personal identity numbers (PINs) for donors that have been diagnosed with malignant lymphoma, search for serum samples harvested at a maximum of ten years before diagnosis."

This use case implies that the system user primarily wants to find samples from specific individuals that

Table 2. BBMRI minimum data set developed during the European BBMRI preparatory phase.

	Attribute	Allowed values	Comment
	Data describing biobanks		
1	BiobankAcronym	ASCII	
2	NameOfBiobank	Free text in English	
3	Institution	Free text in English	
4	URL		Complete http:// address
5	Country	ISO-standard (3166 alpha2), two letter code e.g. SE	
6	ContactName	Free text in English	
7	ContactData	Free text in English	Address, Phone, Mail
	Data describing studies		
8	NameOfStudy	Free text in any language	
9	EnglishStudyName	Free text in English	Translation of study name in English
10	ContactName	Free text in English	
11	ContactData	Free text in English	Address, Phone (E.164, No. 905 – 1.IV.2008), e.g., +46 8 524 877 59, Mail
12	KindOfStudy	Population-based, specific-disease, broad-spectrum of diseases	If "specific-disease", note ICD10
13	CategoriesOfDataCollected	[ClinicalDataAvailable, Diagnosis, Health information, Physiological/biochemical measures, Sociodemographic char., Socioeconomic char., Life habits/Behaviour, Physical environment]	Can be several values.
	Data describing subjects/cases/samples within biobanks		
14	Gender	Male, Female, Other	Gender of subject
15	AgeGroup	Interval [a,b], a>0, b<200, b≥a	a and b should be selected so that k-anonymity is guaranteed. Age group of donor at time for sample collection, number of age groups determined by biobank
16	SampleType	DNA, cDNA/RNA, whole blood, blood cells isolates, serum, plasma, fluids, tissues cryo, tissues paraffin-inbedded, cell-lines	
17	SampleDate	ISO-standard (8601) time format	Date when sample was harvested
18	ClinicalDataAvailable	Yes/No	There exists clinical data related to the sample.
19	OrganCategory	From the BBMRI Detailed descr bio samples	Snomed CT, ISBT 128 etc. Or from the BBMRI Detailed descr bio samples
20	OmicsDataAvailable	Yes/No	Genomics, proteomics etc.
21	RestrictionsOnSampleUse	None, Consent participant, IRB approval, Approval of owner of collection	Can be several values

have already been identified with PINs with the help of e.g. health data registers. In this particular case, the researcher would have found people diagnosed with malignant lymphoma from Swedish cancer registers and via their PINs the researcher would be able to see if they have any samples donated before diagnosis and stored in a biobank. The same information needs to be stored as in use case A.

Matching data sets with the pan-European register BBMRI

Within the BBMRI EU project a minimum data set consisting of 21 attributes describing biobanks and their objects has been proposed [10]. The attributes in table 1 for extending the Swedish Biobank Register Information System for BBMRI.se researchers, can easily be mapped to the 21 attributes in the BBMRI.EU minimal data set described in table 2. Moreover the BBMRI.EU minimal data set is similar to the OECD minimum and recommended data sets for DNA, tissues and isolated cells as described in Appendix 2 of the OECD Best Practice Guidelines for BRCs from 2007 [11].

DISCUSSION

By itself, a biobank can be very useful for many types of studies. However, the power of biobank research will increase significantly if multiple biobanks are connected to enable sharing of information. The move

toward a universal information e-infrastructure for biobanking is directly connected to the issues of semantic interoperability through harmonized services and common ontologies.

In order to achieve a national register for research on biobank samples, it is important to single out potential road blocks early in the development process. Firstly, it is important to make the functionality for BBMRI.se end-user friendly by creating an intuitive graphical user interface (GUI) that requires a minimum amount of time spent learning the system, as success will partly be determined based on how many researchers are using the system. Secondly, biobanking raises ethical questions regarding how to address donor integrity and privacy issues, both regarding what information is stored about a donor and how the data are accessed in an information system. The majority of researchers would agree that biological material without related information has none or very limited value in e.g. epidemiological research. However, there is less agreement on what information can be stored about samples and donors in an information system such as the Swedish Biobank Register with respect to personal integrity. This is the case even when the data are pseudonymized and presented in an aggregated format back to the user. In extending the SBR-IS for research, only aggregated data and data conforming to *k*-anonymity [12] will be presented to the system user. However, this requires that adequate functionality is developed in order to prevent searches breaching the *k*-anonymity rule as well as to ensure that the content is complete and correctly categorized.

BBMRI.se is the first national BBMRI hub to implement the BBMRI minimum data set. It is our belief that this will pave the way for connecting biobanks at the Nordic and European level, contributing to more collaboration between researchers and biobanks. Optimizing usage of biobank samples will not only reduce time and cost, it will produce results of higher statistical significance and in the long term, it will benefit patients and sample donors as new results emerge.

ACKNOWLEDGEMENTS

We would like to thank everybody involved in the collaboration between SBR and BBMRI.se at the Swedish Association of Local Authorities and Regions (SALAR) and Inera AB. We would also like to give a big thank you to the Swedish Research Council for granting the BBMRI.se project (no. 829-2009-6285) as well as to the people involved in the European BBMRI preparatory phase. Finally, thank you Miriam Elfström for linguistic revision of the text.

REFERENCES

1. Pukkala E. Nordic biological specimen banks as basis for studies of cancer causes and control – more than 2 million sample donors, 25 million person years and 100 000 prospective cancers. *Acta Ontol* 2007; **46** (3): 286-307.
2. Silberman S. Libraries of flesh: The sorry state of human tissue storage. *Wired* 2010 (http://www.wired.com/magazine/2010/05/ff_biobanks/).
3. Yuille M, et al. Biobanking for Europe. *Brief Bioinform* 2007; **9** (1): 14-24.
4. The National Board of Health and Welfare, Sweden.
5. Biobanks in Medical Care Act (2002:297) http://www.sweden.gov.se/sb/d/574/a/23126 (Accessed May 12, 2011).
6. www.biobanksverige.se (Accessed May 11, 2011).
7. Civil Registry Act (1991:481) https://lagen.nu/1991:481#P18 (Accessed May 16, 2011).
8. http://www.bbmri.se/en/ (Accessed May 9, 2011).
9. Betsou F. Standard preanalytical coding for biospecimens: Defining the sample PREanalytical code. *Cancer Epidemiol Biomarkers Prev* 2010; **19** (4): 1004-1011.
10. http://www.bbmri.eu/index.php/publications-a-reports Annex 16: Final Report from WP5. (Accessed May 17, 2011).
11. http://www.oecd.org/dataoecd/7/13/38777417.pdf (Accessed May 17, 2011).
12. Sweeney L. *k*-anonymity: a model for protecting privacy. *International journal on Uncertainty, Fuzziness and Knowledge-based Systems* 2002; **10** (5): 557-570.