The Biobanking Research Infrastructure BBMRI_CZ: a Critical Tool to Enhance Translational Cancer Research

Infrastruktura výzkumných biobank BBMRI_CZ: klíčový nástroj translačního výzkumu v onkologii

Holub P.[1], Greplova K.[2], Knoflickova D.[3], Nenutil R.[3], Valik D.[2]

[1] CERIT-SC, Institute of Computer Science, Masaryk University, Brno, Czech Republic, and Masaryk Memorial Cancer Institute, Regional Centre for Applied Molecular Oncology, Brno, Czech Republic
[2] Regional Centre for Applied Molecular Oncology and Department of Laboratory Medicine, Masaryk Memorial Cancer Institute, Brno, Czech Republic
[3] Regional Centre for Applied Molecular Oncology and Department of Pathology, Masaryk Memorial Cancer Institute, Brno, Czech Republic

Summary
We introduce the national research biobanking infrastructure, BBMRI_CZ. The infrastructure has been founded by the Ministry of Education and became a partner of the European biobanking infrastructure BBMRI.eu. It is designed as a network of individual biobanks where each biobank stores samples obtained from associated healthcare providers. The biobanks comprise long term storage (various types of tissues classified by diagnosis, serum at surgery, genomic DNA and RNA) and short term storage (longitudinally sampled patient sera). We discuss the operation workflow of the infrastructure that needs to be the distributed system: transfer of the samples to the biobank needs to be accompanied by extraction of data from the hospital information systems and this data must be stored in a central index serving mainly for sample lookup. Since BBMRI_CZ is designed solely for research purposes, the data is anonymised prior to their integration into the central BBMRI_CZ index. The index is then available for registered researchers to seek for samples of interest and to request the samples from biobank managers. The paper provides an overview of the structure of data stored in the index. We also discuss monitoring system for the biobanks, incorporated to ensure quality of the stored samples.

Key words
biobanking – databases – cancer research

Souhrn
V tomto sdělení popisujeme národní infrastrukturu výzkumných biobank BBMRI_CZ. Infrastruktura byla založena Ministerstvem školství, mládeže a tělovýchovy a stala se partnerem evropské infrastruktury biobank BBMRI. Infrastruktura je navržena jako síť biobank, které skladují vzorky získané od asociovaných zdravotnických institucí. Biobanky sestávají z dlouhodobého úložiště (různé typy tkání klasifikované podle diagnózy, peroperační sérum, genomová DNA, RNA) a krátkodobého úložiště (séra pacientů odebíraná v čase). Diskutujeme způsob práce infrastruktury, který musí odpovídat její distribuované povaze: získávání vzorků musí být doprovázeno extrakcí dat z nemocničních informačních systémů a tato data musejí být katalogizována v centrálním indexu pro potřeby vyhledávání. Jelikož BBMRI_CZ slouží pouze pro potřeby vědy a výzkumu, jsou data před uložením do indexu anonymizována. Index je poté k dispozici registrovaným výzkumným pracovníkům, kteří mohou o vybrané vzorky podat žádosti správcům biobank. Článek poskytuje přehled struktury dat uložených v indexu. Diskutujeme také monitorovací systém biobank, který je do BBMRI_CZ začleněn pro dohled nad dodržováním kvality uskladnění vzorků.

Klíčová slova
biobanka – databáze – výzkum rakoviny

This study was supported by the European Regional Development Fund and the State Budget of the Czech Republic (RECAMO, CZ.1.05/2.1.00/03.0101) and by Large Infrastructure Projects of Czech Ministry of Education BBMRI_CZ LM2010004 and CERIT-SC CZ.1.05/3.2.00/08.0144.

Práce byla podpořena Evropským fondem pro regionální rozvoj a státním rozpočtem České republiky (OP VaVpI – RECAMO, CZ.1.05/2.1.00/03.0101) a projekty Velkých infrastruktur pro VaVaI MŠMT BBMRI_CZ LM2010004 a CERIT-SC CZ.1.05/3.2.00/08.0144.

The authors declare they have no potential conflicts of interest concerning drugs, products, or services used in the study.

Autoři deklarují, že v souvislosti s předmětem studie nemají žádné komerční zájmy.

The Editorial Board declares that the manuscript met the ICMJE "uniform requirements" for biomedical papers.

Redakční rada potvrzuje, že rukopis práce splnil ICMJE kritéria pro publikace zasílané do biomedicínských časopisů.

Dalibor Valik, M.D., Ph.D., DABCC, FACB, Assoc. Prof.
Masaryk Memorial Cancer Institute
Regional Centre for Applied Molecular Oncology and Department of Laboratory Medicine
Zluty kopec 7
656 53 Brno
Czech Republic
e-mail: valik@mou.cz

Submitted/Obdrženo: 12. 10. 2012
Accepted/Přijato: 25. 10. 2012

Introduction

The BBMRI_CZ, Czech national research biobanking infrastructure, was founded by the Ministry of Education to set up a network of biobanks for cancer research in the Czech Republic. This activity is coordinated by the National coordinating node, Masaryk Memorial Cancer Institute (MMCI), with further biobanking units affiliated to faculties of medicine. Biobanks collect and store biological material from cancer patients for the long term that would be otherwise lost for future biomedical research. The general concept of the biobank infrastructure is summarised and approved by the Government in the text of the National Roadmap of Large Infrastructures under the Paragraph 2.5.3. Priority projects: *BANK OF CLINICAL SAMPLES (BBMRI_CZ), The biobank of clinical samples is an existing large infrastructure founded and maintained by MMCI and is functionally bound to the OP R&DI RECAMO project. The Biobank of clinical samples at MMCI was certified by the management board of BBMRI as an associated organisation and became the coordinator of the Czech part of the pan-European research infrastructure BBMRI (Biobanking and biomolecular resources research infrastructure) under BBMRI_CZ name [1].*

The BBMRI_CZ, a Czech national research biobanking infrastructure, is a unique system comprising long-term and short-term storage of biological samples. It is a distributed system spanning several institutions, from healthcare institutions being sample providers, through universities as biobank maintainers, to researchers coming from various institutions that may request samples from a biobank. The infrastructure deals with the patient data and thus it needs specific approaches to design information technology (IT) infrastructures to index and protect data describing stored samples and to make the indices available to the researchers both throughout the Czech Republic and within the European BBMRI.eu project in the future.

This project is a part of pan-European biobanking project called BBMRI.eu [2]. The data storage and access is a part of several workpackages of the BBMRI.eu project as shown by a concept of generic BBMRI Catalogue service [3]. Our design generally follows the identifier specifications suggested by the BBMRI D5.2 Deliverable [4]. Because of the unavailability of implementations that could be taken over and the specifics of BBMRI_CZ biobank structure, as well as legal requirements in the Czech Republic, the BBMRI_CZ project decided to implement a custom interim data management infrastructure. A reference data gathering infrastructure is represented by the National Oncology Register (NOR) [5] of the Czech Republic, which collects information about treatment of patients suffering from any type of malignancy. The data model proposed for the BBMRI_CZ biobank is designed to be consistent with the NOR and later research will be focused on their mutual synergy, given the legislative restrictions imposed on patient data handling.

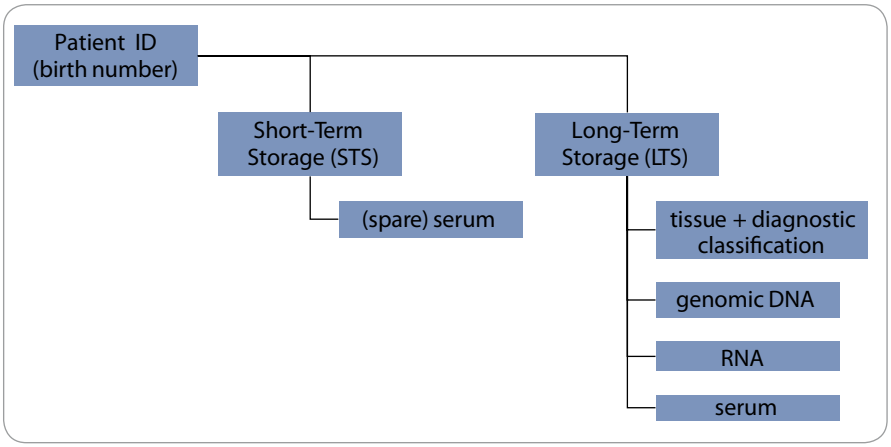

Fig. 1. Structure of the BBMRI_CZ biobank.

Structure of the Biobank and Operational Workflow in BBMRI_CZ

All the samples collected within the biobank are tied to the identity of the patient (i.e., so called „birth number" augmented with a disambiguating extension to resolve erroneous birth number duplicates). As shown in Fig. 1, the Biobank comprises two major components: a long term storage (LTS) repository and a short term storage (STS) repository. The LTS repository collects various types of tissues (tumour, metastases, non-tumour) classified by diagnosis, serum at surgery, genomic DNA and RNA. This part of the biobank is filled with low frequency, typically at the moment of the patient's primary surgery. Short term storage contains sera only and is iteratively updated at each patient visit to the hospital when the blood specimen is taken for the determination of tumour markers. The short term storage serum repository thus stores leftovers of tumour marker patient material for a period of up to one year. The design uniqueness of the BBMRI_CZ thus stems from the fact that the "biobanking entity" is the patient rather than just a type of the material stored. An overview of storage operations in LTS and STS parts of the Biobanks related to patient treatment is shown in Fig. 2.

From the organisational perspective, the samples are provided by the healthcare institutions, where patients are treated, together with selected subset of the patient's clinical data. Samples are hosted by one of the participating Biobanks, which is typically operated by a university medical school distinct from the healthcare institution. A notable exception is represented by Masaryk Memorial Cancer Institute (MMCI), which plays a role as both sample provider and Biobank itself, being the largest comprehensive cancer centre at the national level. The overall operation of the Biobanks is governed and coordinated by MMCI, which also maintains the central data infrastructure.

User Access

At present the primary purpose of a Biobank is to serve the needs of the resear-

chers at the respective location and elsewhere. Before accessing the Biobank, each researcher needs to be associated with a project that is approved by at least one of the participating institutions possessing a BBMRI_CZ-associated Biobank. Identity of the users is verified using so-called nationalised authentication mechanisms, typically eduID.cz [6] in the Czech Republic. Access of registered researchers to the Biobank is twofold: (i) the user interacts with the central index to look up samples based on data available in the Biobank indices, (ii) the user requests specific sample sets from the participating Biobanks. Each sample request is approved or denied by the manager of the Biobank containing the sample. All the operations are logged and decisions about the requests are registered.

Monitoring of Biobanks

Another important part of the BBMRI_CZ infrastructure is long-term monitoring of physical parameters of storage containers in order to ensure quality of stored samples and to monitor remaining free space. Physical parameters of each Biobank are continuously measured and stored locally; in the future they will be also transferred to the central BBMRI_CZ infrastructure. Should the operation limits of a Biobank be exceeded, both Biobank operator and the infrastructure coordinator are automatically notified.

Data Structure of BBMRI_CZ

All the samples and linked data are bound to the identity of a patient. Being designed solely for research purposes, the Biobank infrastructure ensures anonymisation of the data as a part of the export process from a hospital information system (HIS). There are two basic requirements on the anonymisation process: (a) identification of samples that belong to the same patient, albeit being stored in different biobanks (because the patient was treated at two separate healthcare institutions), and (b) distinguishing samples that belong to different patients. The system is designed to work with either internal or external anonymisation. The internal anonymisation is based on application of a unidirectional cryptographic function applied on the patient's identifier – currently SHA512 is applied to the patient birth number with disambiguating extension augmented with a random string generated specifically for each periodic export of the data. The augmenting random string is automatically discarded immediately after the export. This approach mitigates problems with brute-force attacks based on listing of all the birth numbers but it also introduces transient inconsistency: while each export set from all the Biobanks is consistent, identifiers change between successive exports. The external anonymisation assumes offloading the process to a trusted external entity (such as KSRZIS). This process maintains temporary consistency, but increases cost of operation of the infrastructure.

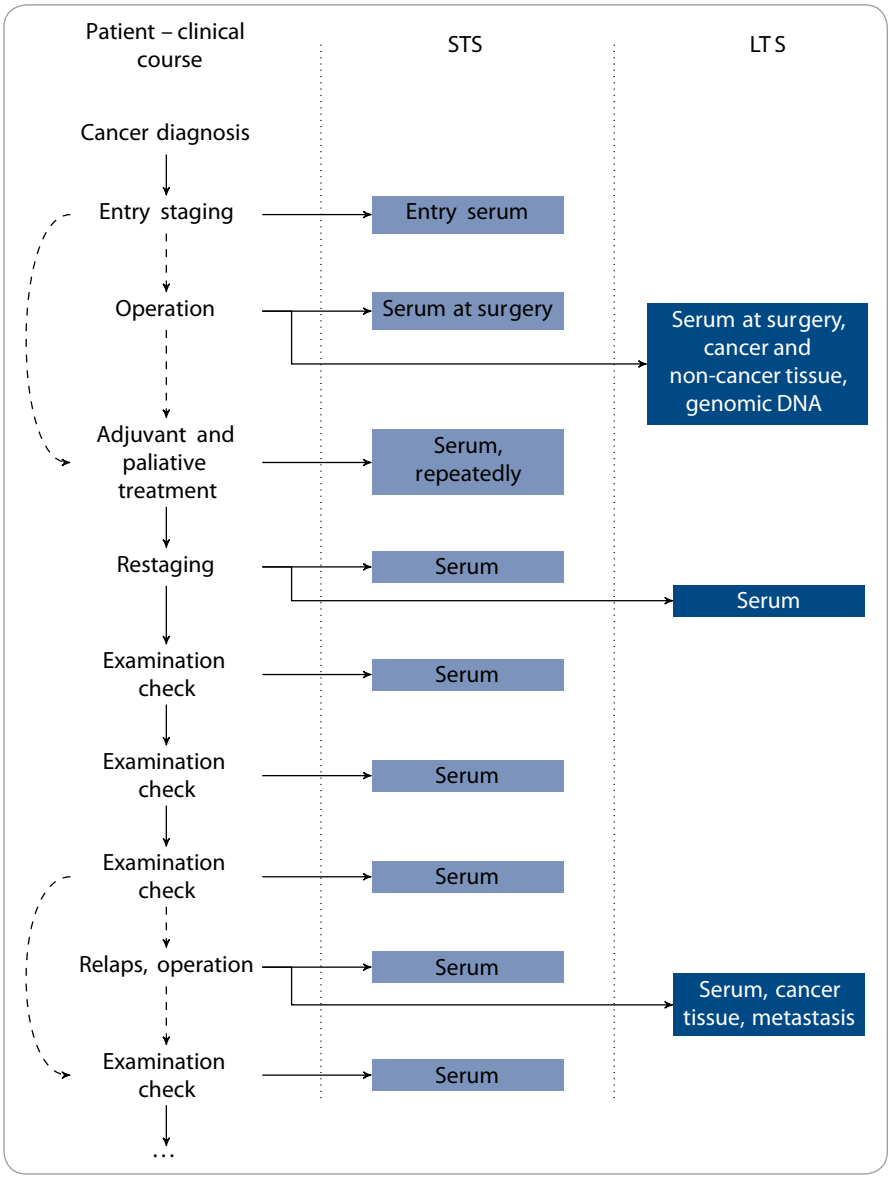

Fig. 2. Storage operations in the biobank.

Data maintained about the samples is aggregated into four modules: tissue module, serum module, genomic module and short term serum module. For each of the modules, the following data is stored:

Tissue module

- sample identifier (composed of a Biobank identifier, year, and the sequential number of the sample within each year)

- type of the tissue (tumour, metastasis, benign tissue, non-tumour tissue)
- total number of samples stored (aliquots)
- number of samples available in the Biobank (a certain amount of the material may be reserved for reference and verification purposes)
- TNM classification
- pTNM classification
- grading
- date and time of termination of vascular supply
- date and time of sample freezing

Serum module

- sample identifier (composed of a Biobank identifier, year, and the sequential number of the sample within each year)
- total number of samples stored (aliquots)
- number of samples available from the biobank (a certain amount of material may be reserved for reference and verification purposes)
- date and time of sample taking

Genomic module

- sample identifier (composed of a Biobank identifier, year, and the sequential number of the sample within each year)
- type of the sample (gD – genomic DNA, PK – full blood)
- total number of samples stored (aliquots)
- number of samples available from the biobank (a certain amount may be reserved for reference and verification purposes)
- date and time of sample taking

Short term serum repository

- sample identifier (composed of a Biobank identifier, year, and the sequential number of the sample within each year)
- diagnosis
- date and time of sample taking

For each patient, the Biobank also stores patient's informed consent that his/her data and samples may be used for research purposes.

Data Protection and Access Control

As discussed above, the identifiers of patients are anonymised – this should provide sufficient protection unless the system deals with rare diseases. Since BBMRI_CZ infrastructure is not supposed to store samples of rare diseases, we consider the proposed approach appropriate for identity protection. For later extension to rare diseases, the system may be enhanced with some of the k-anonymisation approaches [7].

Conclusions

Establishing a networked system of cancer research-focused Biobanks affiliated to academic institution is a challenging endeavour. The unique design of storing not only the tissue material but also longitudinal strings of sera enables access to patient-derived material during the course of the complex patient treatment, thus reflecting pathophysiological and treatment-induced changes in the course of the disease. Designed this way, the research Biobanks will become truly critical tools to enhance translational cancer research.

References

1. Msmt.cz [online]. Ministerstvo školství mládeže a tělovýchovy; cMŠMT 2006-2012 [upd. květen 2011; cit. 22. října 2012]. Dostupné z: http://www.msmt.cz/mezinarodni-vztahy/vyzkum-a-vyvoj-1/aktualizovana-cestovni-mapa-cr-velkych-infrastruktur-pro?highlightWords=cestovn%C3%AD+mapa.
2. BBMRI.eu [online]. Biobanking and Biomolecular Resources Research Infrastructure; cBBMRI [cit. 2012 October 22]. Available from: http://www.bbmri.eu.
3. BBMRI.eu [online]. Wichmann HE, Kuhn KA, Waldenberger M et al. BBMRI Catalogue – System and Data Architecture. Technische Universität München, 2010; cBBMRI [upd. 4 October 2010; cit. 2012 October 22]. Available from: http://www.bbmri.eu/bbmri/index.php?option==com_docman&task=doc_details&gid=263&ltemid=97.
4. BBMRI.eu [online]. Kuhn KA. D5.2 Strategy for Unique and Secure Identities for Specimens, Subjects, and Biobanks. BBMRI, 2009. cBBMRI [upd. 14 April 2011; cit. 2012 October 22]. Available from: http://www.bbmri.eu/bbmri/index.php?option=com_docman&task=doc_details&gid=311&ltemid=97.
5. Svod.cz [online]. Dušek L, Mužík J, Kubásek M, et al. Epidemiologie zhoubných nádorů v České republice. Masarykova univerzita 2005. [cit. 25. října 2012]. Dostupné z: http://www.svod.cz.
6. Sova M, Tomášek J. SAML Metadata Management for eduID.cz. In: Proceedings, CESNET Conference 2008 – Security, Middleware, and Virtualization, Praha, CESNET z.s.p.o., 2008: 23–28.
7. Ciriani V, De Capitani di Vimercati S, Foresti S et al. k-Anonymity. In: Yu T, Jajodia S (eds). Secure Data Management in Decentralized Systems, Advances in Information Security. Vol. 33. Springer 2007: 323–353.