DATABASE
The Journal of Biological Databases and Curation

Database, 2019, 1–5
doi: 10.1093/database/baz105
Database tool

Database tool

BioSCOOP – Biobank Sample Communication Protocol. New approach for the transfer of information between biobanks

J. Jarczak[1,2], J. Lach[1,2], P. Borówka[1,3], M. Gałka[4], M. Bućko[5], B. Marciniak[1,2] and D. Strapagiel[1,2,*]

[1]Biobank Lab, Department of Molecular Biophysics, Faculty of Biology and Environmental Protection, University of Łódź, Łódź, Poland, [2]BBMRI.pl Consortium, Wrocław, Poland, [3]Department of Anthropology, Faculty of Biology and Environmental Protection, University of Łódź, Łódź, Poland, [4]Independent IT Specialist and [5]Bee2code sp. z o.o., ul. Daszyńskiego 5; 44-100 Gliwice

*Corresponding author: Biobank Lab, Department of Molecular Biophysics, Faculty of Biology and Environmental Protection, University of Łódź, ul. Pilarskiego 14/16, 90-237, Łódź, Poland; e-mail: dominik.strapagiel@biol.uni.lodz.pl

Citation details: Jarczak,J., Lach,J., Borówka,P. *et al.* BioSCOOP – Biobank Sample Communication Protocol. New approach for the transfer of information between biobanks. *Database* (2019) Vol. 2019: article ID baz105; doi:10.1093/database/baz105

Received 17 April 2019; Revised 6 June 2019; Accepted 2 August 2019

Abstract

Dynamic development of biobanking industry (both business and science) resulted in an increased number of IT systems for samples and data management. The most difficult and complicated case for the biobanking community was cooperation between institutions, equipped with different IT systems, in the field of scientific research, mainly data interchange and information flow. Tools available on the market relate mainly to the biobank or collection level. Efficient and universal protocols including the detailed information about the donor and the sample are still very limited. Here, we have developed BioSCOOP, a communication protocol in the form of a well documented JSON API. The main aim of this study was to harmonize and standardize the rules of communication between biobanks on the level of information about the donor together with information about the sample. The purpose was to create a communication protocol for two applications: to transfer the information between different biobanks and to allow the searching and presentation of the sample and data sets.

Introduction

Every organization, sooner or later will have to face the problem of difficulties in communication. A precise information flow is crucial and allows for quick decision making, prevents conflicts, and facilitates daily work. The lack of communication hinders current work and leads to wasting time. In the world of biobanks, information flow and data transfer are the basis of efficient functioning

of units which were created to collect biological material for further advanced research. Dynamic development of biobanking industry (both business and science) resulted in an increased number of IT systems for samples and data management. The most difficult and complicated case for biobanking community was the cooperation between institutions, equipped with different IT systems, especially in the field of data harmonisation (data interchange and information flow). The problem starts when we want to describe parameter when different scales are commonly used. Temperature can be an excellent example. It can be measured in at least three different scales (Kelvin, Fahrenheit and Celsius), which refer to the same value but are described using three different units. Exchange of information about temperature between two entities using different scales can lead to misunderstanding. Biobanks are invaluable sources of data with huge potential of biological material reuse, sometimes limited by communication restrictions resulted from heterogeneity of biorepositories (1–3). In their repositories, they store, sample sets that are supplemented by the large data sets composed of (depending on the type of biobank): the list of phenotypic features of the donor, information about the diseases and all medical history, lifestyle information, information about the sample, information about storage and quality parameters etc. For researchers, any information about the collected samples and data is extremely important in connection to the data they generate. The need to conduct research on a very specific and precisely defined sample set is undeniable. In the era of big data and increasing importance of personalized medicine, data visibility and access, storage, management and integration has become a major issue in biobanking and biomedical research (4). Increasing number of specialized biorepositories and expansion of available data types produced by biomedical or research centres, require adequate sample information and management systems (e.g. BIMS, BBMS, LIMS) for location and integration of metadata, along with well-defined sample description standards for stored biological material (e.g. ICD-10, SPREC, BRISQ) (5–9). Due to varying specific features of biomedical facilities or biobanks worldwide, IT solutions for sample information and management are often tailor-made for biorepositories, stemming directly from the type of basic research, used biological material, storage requirements or survey restrictions (5). These internal standards become a challenge for biobank- to biobank communication or data exchange throughout biobanking networks, primarily created for facilitation of data interchange. Direct communication between biobanks, which are providers of biological material for secondary research (in accordance with the Tri-Council Policy Statement (10)) is troublesome due to tree key limitations: divergent description of the samples, different levels of accuracy about the donor and incompatible IT solutions for data storage and transfer (10,11). Currently, there are many standards for sample management implemented by biorepositories which touch upon issues of donor-sample description (12–14), sample SOPs (15–17), directories of biological material collections (16,18), ontology of collections and biobanks (18–29), network and integration protocols for biobanks (30–36), or even biobank-biobank matching algorithms (37). These factors reflect the complexity of unification of individual sample description for communication/exchange protocol between biobanks. Universal FOSS (free and open source software) or protocols containing minimal information about sample-donor operating on common communication IT infrastructure, are still very limited. However, there are attempts to improve communication between biobanks and first solutions to facilitate sample location and access such as service Negotiator 1.0 made by BBMRI-ERIC (38) and Sample Request Portal (open source portal PODIUM) prepared by BBMRI-nl (national node of BBMRI-ERIC in Netherlanden). The proposed standard joins different ontologies used in sample-donor description models such as MIABIS, BRISQ etc., use recognised disease ontologies e.g. ICD-10, with parameters used in ergonomics, anthropometry and biomechanics e.g. ISO-TC159/SC3: therefore, it collectively provides effective networking and resource sharing between biobanks. The main aim of this study was to harmonize and standardize the rules of communication between biobanks on the level of information about the donor. To address these issues, BioSCOOP was created as a communication protocol for two applications: to transfer the information between different biobanks and to allow searching and presentation of sample and data sets.

Results

BioSCOOP has the form of a well documented JSON API which describes an organized data format for a list of attributes describing the donor with particular emphasis on the phenotype, anthropological measurements, medical data and sample material. The software application of this standard was created using Swagger Editor, a tool for API creation, to be compliant with Open API Specification. BioSCOOP has been deposed on Github, as YAML file and can be easily imported into Swagger Editor or any other text editor as a described JSON.

Furthermore, an exemplary data set has also been prepared. It can be downloaded and used for test sample search using the proposed browser – Bioface. It was provided to guide users through sample search based on BioSCOOP standard.

The list of features included in BioSCOOP is listed in Table S1 (supplementary information).

Implementation

Import of data in BioSCOOP format has been implemented in the related project, Bioface. Bioface has a distributed architecture and is designed as a browser for the members of Polish Biobanking Network (PBN) (39) as well as a broader group of biobanks and researchers, in order to search for samples from different biobanks and biorepositories. It is a part of IT infrastructure for PBN, which includes both central and distributed solutions for data collection and sharing. Implementation was divided into three independent steps:

1. Test data set preparation – an exemplary data set was prepared using Microsoft Excel spreadsheet. It contains randomly generated information about 200 database records mimicking samples collected from 200 mock donors. The provided information includes: birth date, place of birth and residence, sex, ethnic origin, skin tone, hair and eye colour, blood group, parameters like WHR (waist hip ratio), BMI (body mass index), CI (Corpulence index), some of anthropological features, diseases and medical procedures undergone by the donor and form of sample material (Tab. S1.). This information was supplemented by donor ID, collection ID, sample ID and measurement/event date timestamp. The format includes also information on the source of included data (donor questionnaire). This test file was initially prepared in.csv format. Then, the data set has been transformed with the use of a homemade script written in Python. This script, by imported data from.csv file, and converted them into JSON, according to the data format written in BioSCOOP.
2. Registration in Bioface – this step was necessary to carry out the sample search procedure. For testing purposes, we first created a dummy account with a mock biobank in Bioface. We subsequently uploaded the previously generated JSON-format data set and used it to perform test searches of the included mock samples.
3. Sample search – various queries have been tested to obtain defined sample set. Queries structure is characteristic for Apache Solr search platform which is a base of Bioface.

   Examples of basic queries structure:
   a. Basic queries:
   field_name:value; e.g. gender:male
   b. Phrase query:
   field_name:“string value”; e.g. birthPlace:“Gdansk, Poland”
   c. Range query:
   numeric_field_name:[lower_limit TO upper_limit]; e.g. bmi:[18 TO 23]

   Also using logical operators to combine subsequent parts of query is posible. The above examples do not exhaust the possibility of creating queries in the used engine, which are described in more detail in the Apache Solr documentation (https://lucene.apache.org/solr/guide/).

Conclusions and future developments

BioSCOOP was created as a communication protocol and aims to facilitate and improve the information transfer in a large network of biobanks. The members of the Polish Biobanking Network will be involved in first implementation of described protocol. On the basis of this, there are further goals such as gathering specialists in many fields of science in one workgroup to create the most accurate way for description data collected by biobanks and scientists. We discuss also future developments. The next step is implementation of BioSCOOP in the BIMS system, currently being created by the Polish Biobanking Network. BioSCOOP will also be used as a data import format in data processing IT software developed by the Polish Biobanking Network.

Acknowledgements

We thank Łukasz Pułaski for a language correction and a thorough review of the manuscript. Funding: The work was supported by Polish Ministry of Science and Higher Education no. DIR/WK/2017/01:“Biobank network in Poland, within the BBMRI-ERIC Research Infrastructure of Biobanks and Biomolecular Resources” and by the Polish POPC Grant 02.03.01-00-0012/17-00 from the European Regional Development Fund.

Availability: BioSCOOP is available under open source licence: https://github.com/BiobankLab/BioSCOOP

Contact: Biobank Lab, Department of Molecular Biophysics, Faculty of Biology and Environmental Protection, University of Łódź, ul. Pilarskiego 14/16, 90-237, Łódź, Poland; e-mail: dominik.strapagiel@biol.uni.lodz.pl

Supplementary data

Supplementary data are available at *Database* Online.

- Bioscoop available from: http: https://github.com/BiobankLab/BioSCOOP
- Table of features describing the content of Bioscoop – Tab. S1
- Exemplary data set to perform an usability check – file in JSON format https://github.com/BiobankLab/BioSCOOP
- Demo Bioface for testing purpose - available from: http://biobank.uni.lodz.pl:8082/bioface/

Conflict of interest. None declared.

References

1. Paskal,W., Paskal,A.M., Debski,T. *et al.* (2018) Aspects of modern biobank activity - comprehensive review. *Pathol Oncol Res*, **24**, 771–785.
2. Holzinger,A., and Huppertz,B. (2014) Biobanks – a source of large biological data sets: open problems and future challenges. *Interactive Knowledgle Discovery and Data Mining in Biomedical Informatics*: State-of-the-Art and Future Challenges. Lecture Notes in Computer Science. Heidelberg, Berlin, Springer. 317–330. https://doi.org/10.1007/978-3-662-43968-5_18.
3. Müller,H., Rheis,R., Zatloukal,K. *et al.* (2015) State-of-the-Art and Future Challenges in the Integration of Biobank Catalogues. In: *Smart Health*, *Lecture Notes in Computer Science* **8700**, Heidelberg, Springer. 261–273. https://doi.org/10.1007/978-3-319-16226-3_11.
4. Fransson,M.N., Rial-Sebbag,E., Brochhausen,M. *et al.* (2015) Toward a common language for biobanking. *Eur J Hum Genet*, **23**, 22–28.
5. Smith,B., Arabandi,S., Brochhausen,M. *et al.* (2015) Biomedical imaging ontologies: a survey and proposal for future work. *J Pathol Inform*, **6**, 37.
6. Moore,H.M., Kelly,A.B., Jewell,S.D. *et al.* (2011) Biospecimen reporting for improved study quality (BRISQ). *J Proteome Res*, **10**, 3429–3438.
7. Bendou,H., Sizani,L., Reid,T. *et al.* (2017) Baobab laboratory information management system: development of an open-source laboratory information management system for biobanking. *Biopreserv Biobank*, **15**, 116–120.
8. Michalska-Madej,J., Dobrowolska,S. and Strapagiel,D. (2018) Application of BBMS in the biobanks storage management. *Eur J Transl Clin Med*, **1**, (Suppl.4): 72.
9. Dobrowolska,S., Michalska-Madej,J., Słomka,M. *et al.* (2019) Biobank Łódź® – population based biobank at the University of Łódź, Poland. *Eur J Transl Clin Med*, **2**, 85–95.
10. Olund,G., Lindqvist,P. and Litton,J.E. (2007) BIMS: an information management system for biobanking in the 21st century. *Ibm Systems Journal*, **46**, 171–182.
11. (2014) Tri-Council, Canadian. Tri-Council Policy Statement 2 — the latest edition of Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans.
12. Tasse,A.M. (2016) A comparative analysis of the legal and bioethical frameworks governing the secondary use of data for research purposes. *Biopreserv Biobank*, **14**, 207–216.
13. (1991) National Center for Health Statistics (US). International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM).
14. (2004) World Health Organization. ICD-10: international statistical classification of diseases and related health problems: tenth revision, 2nd ed.
15. El-Sappagh,S., Franda,F., Ali,F. *et al.* (2018) SNOMED CT standard ontology based on the ontology for general medical science. *Bmc Medical Informatics and Decision Making*, **18**.
16. Kuhn K, B.R., Spengler H (2015) *BBMRI catalogue. Journal of Clinical Bioinformatics*, **5** (1), S19.
17. Wichmann,H.E., Kuhn,K.A., Waldenberger,M. *et al.* (2011) Comprehensive catalog of European biobanks. *Nat Biotechnol*, **29**, 795–797.
18. Holub,P., Swertz,M., Reihs,R. *et al.* (2016) BBMRI-ERIC directory: 515 biobanks with over 60 million biological samples. *Biopreserv Biobank*, **14**, 559–562.
19. Brochhausen,M., Fransson,M.N., Kanaskar,N.V. *et al.* (2013) Developing a semantically rich ontology for the biobank-administration domain. *J Biomed Semantics*, **4**, 23.
20. Norlin,L., Fransson,M.N., Eriksson,M. *et al.* (2012) A minimum data set for sharing biobank samples, information, and data: MIABIS. *Biopreserv Biobank*, **10**, 343–348.
21. Merino-Martinez,R., Norlin,L., van Enckevort,D. *et al.* (2016) Toward global biobank integration by implementation of the minimum information about BIobank data sharing (MIABIS 2.0 Core). *Biopreserv Biobank*, **14**, 298–306.
22. Segagni,D., Tibollo,V., Dagliati,A. *et al.* (2011) The ONCO-I2b2 project: integrating biobank information and clinical data to support translational research in oncology. *Stud Health Technol Inform*, **169**, 887–891.
23. Gremse,M., Chang,A., Schomburg,I. *et al.* (2011) The BRENDA tissue ontology (BTO): the first all-integrating ontology of all organisms for enzyme sources. *Nucleic Acids Res*, **39**, D507–D513.
24. Lehmann,S., Guadagni,F., Moore,H. *et al.* (2012) Standard pre-analytical coding for biospecimens: review and implementation of the sample PREanalytical code (SPREC). *Biopreserv Biobank*, **10**, 366–374.
25. Forrey,A.W., McDonald,C.J., DeMoor,G. *et al.* (1996) Logical observation identifier names and codes (LOINC) database: a public use set of codes and names for electronic reporting of clinical laboratory test results. *Clin Chem*, **42**, 81–90.
26. Stoeckert, C.J., Jr., Parkinson, H. (2003) The MGED ontology: a framework for describing functional genomics experiments. *Comp Funct Genomics*, **4**, 127–132.
27. Hochedlinger,N., Nitzlnader,M., Falgenhauer,M. *et al.* (2015) Standardized data sharing in a paediatric oncology research network–a proof-of-concept study. *Stud Health Technol Inform*, **212**, 27–34.
28. Brochhausen,M., Zheng,J., Birtwell,D. *et al.* (2016) OBIB-a novel ontology for biobanking. *J Biomed Semantics*, **7**, 23.
29. Weiler,G., Schroder,C., Schera,F. *et al.* (2014) P-BioSPRE-an information and communication technology framework for transnational biomaterial sharing and access. *Ecancer*, **8**, 401.
30. Hammond,W.E. (1991) Health level 7: an application standard for electronic medical data exchange. *Top Health Rec Manage*, **11**, 59–66.
31. Litton,J.E. (2011) Biobank informatics: connecting genotypes and phenotypes. *Methods Mol Biol*, **675**, 343–361.
32. Izzo,M. (2016) The JSON-Based Data Model. In: *Biomedical Research and Integrated Biobanking: An Innovative Paradigm for Heterogeneous Data Management*. Springer Theses (Recognizing Outstanding Ph.D. Research). Springer, Cham. doi: org/10.1007/978-3-319-31241-5_3.
33. Chen,C., Wulff,R.T., Sholle,E.T. *et al.* (2018) Evaluating generalizability of a biospecimen informatics approach: support for local requirements and best practices. *AMIA Jt Summits Transl Sci Proc*, **2017**, 55–62.
34. Dangl,A., Demiroglu,S.Y., Gaedcke,J. *et al.* (2010) The IT-infrastructure of a biobank for an academic medical center. *Stud Health Technol Inform*, ***2010***; **160**, (Pt 2), 1334–1338.

35. Ferretti,Y., Miyoshi,N.S.B., Silva,W.A. *et al.* (2017) BioBankWarden: a web-based system to support translational cancer research by managing clinical and biomaterial data. *Computers in Biology and Medicine*, **84**, 254–261.
36. Eminaga,O., Semjonow,A., Oezguer,E. *et al.* (2014) An electronic specimen collection protocol Schema (eSCPS). *Methods of Information in Medicine*, **53**, 29–38.
37. Spjuth,O., Krestyaninova,M., Hastings,J. *et al.* (2016) Harmonising and linking biomedical and clinical data across disparate data archives to enable integrative cross-biobank research. *European Journal of Human Genetics*, **24**, 521–528.
38. Pang,C., Kelpin,F., van Enckevort,D. *et al.* (2017) BiobankUniverse: automatic matchmaking between datasets for biobank data discovery and integration. *Bioinformatics*, **33**, 3627–3634.
39. Proynova,R., Alexandre,D., Lablans,M. *et al.* A decentralized IT architecture for locating and negotiating access to biobank samples. *Stud Health Technol Inform*, **243**, 75–79.
40. Witon,M., Strapagiel,D., Glenska-Olender,J. *et al.* (2017) Organization of BBMRI.Pl: the polish biobanking network. *Biopreserv Biobank*, **15**, 264–269.