Advances in Informatics, Management and Technology in Healthcare
J. Mantas et al. (Eds.)

doi:10.3233/SHTI220722

Integration of Biobanking Architecture with Genomics Data: Genomics Integrated Biobanking Ontology (GIBO)

Nitya SINGH[a1], Naomi BRAUN[b], William HOGAN[b] and Mathias BROCHHAUSEN[c]
[a] *Emerging Pathogens Institute, Food Systems Institute, Animal Sciences Department, University of Florida, Gainesville, Florida, USA*
[b] *Department of Health Outcomes & Biomedical Informatics, University of Florida, Gainesville, Florida, USA*
[c] Dept. of Biomedical Informatics, University of Arkansas for Medical Sciences, USA

Abstract. Integration of clinical-pathological information of Biobanks with genomics-epidemiological data/inferences in a structured and consistent manner, mitigating inherent heterogeneities of sites/sources of data/sample collection, processing, and information storage hurdles, is primary to achieving an automated surveillance system. Genomics Integrated Biobanking Ontology (GIBO) presents a solution for preserving the contextual meaning of heterogeneous data, while interlinking different genomics and epidemiological concepts in machine comprehensible format with the biobank framework. GIBO an OWL ontology introduces 84 new classes to integrate genomics data relevant to public health.

Keywords. Genomic epidemiology, biobanks, ontological integration

1. Introduction

Integration of public health data across clinical-pathological, genomic, epidemiological, data storage, and analysis platforms present a great opportunity for real-time tracking of any vector/food/environment-borne infectious disease transmission and their incidence estimates. The heterogeneous nature of data generation and storage systems presents an inherent challenge and only a scarce success has been achieved so far [1]. In national public health setups, epidemiological bureaus maintain clinical-demographic records, case recalls, and interviews information on infectious disease cases, while public health laboratories and biobanks house information, and molecular sequencing facilities store the genomic information which syncs with National surveillance system. Heterogenous architecture/terminology within these systems, causes inconsistencies and redundancies during integration, impeding clear inferences, e.g. temporal patterns and potential outbreaks of causative infectious agents, required for infection prevention. In this paper, we provide a semantically rich representation (usable by Semantic Web Technology, SWT) to interlink data from heterogeneous demographic, clinical, epidemiological, and genomic sources to support an integrated public health surveillance system architecture.

[1] Corresponding author, Nitya Singh, Emerging Pathogens Institute, Food Systems Institute, Animal Sciences Department, University of Florida, Gainesville, Florida, USA; E-mail: nitya11@ufl.edu.

2. Methods

The Genomics Integrated Biobanking Ontology (GIBO) (hosted at Github: https://github.com/Nits11/GIBO) is a publicly available ontology implemented using Web Ontology Language (OWL) [2]. In its development, we followed the OBO Foundry principles [3] and recommendations by Brochhausen *et al.* [4]. GIBO was built using the architectural backbone of OBIB [5] with realism-based ontology development [6] as the general methodology to inform knowledge representation. The consistency of GIBO was checked using HermiT 1.4.3.456 and FaCT 1.6.5++ reasoners.

3. Results

Eighty-four new classes were created to represent the genomics and clinical domains, along with classes from pre-existing ontologies like OBI, OBIB, OMIABIS, OMRSE, GEO which were also used to represent the logical architecture of the different targeted domains. The five most prominent use cases were identified, along with 10 competency questions, and their axiomatic solutions have been developed for the automated knowledge discovery process.

4. Discussion and conclusions

The design of GIBO provides a logical framework to address heterogeneous data integration issues, preventing data loss and hurdles in time-constrained surveillance and outbreak detection. It would support the knowledge discovery by rationalizing answers to crucial clinical and molecular epidemiological questions which are difficult to ask otherwise. As, the next steps, GIBO will be tested with a Triple store for a heterogeneous sample dataset (from sources including epidemiological and clinical databases, biobanks, molecular sequencing facilities, and public genomic databases) to solve competency questions with SPARQL queries for drawing relevant research inferences and knowledge discovery in the public health interest.

References

[1] Griffiths E, Dooley D, Graham M, Domselaar G van, Brinkman FSL, Hsiao WWL. Context is everything: Harmonization of critical food microbiology descriptors and metadata for improved food safety and surveillance. Frontiers in Microbiology. 2017 Jun 26;8:1068.

[2] Hitzler P, Krötzsch M, Parsia B, Patel-Schneider PF, Rudolph S. OWL 2 Web Ontology Language Primer 2012; Second edition [Internet].

[3] Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg LJ, Eilbeck K, Ireland A, Mungall CJ, Leontis N. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nature biotechnology. 2007 Nov;25(11):1251-5.

[4] Brochhausen M, Fransson MN, Kanaskar NV, Eriksson M, Merino-Martinez R, Hall RA, Norlin L, Kjellqvist S, Hortlund M, Topaloglu U, Hogan WR. Developing a semantically rich ontology for the biobank-administration domain. Journal of biomedical semantics. 2013 Dec;4(1):1-9.

[5] Brochhausen M, Zheng J, Birtwell D, Williams H, Masci AM, Ellis HJ, Stoeckert CJ. OBIB-a novel ontology for biobanking. Journal of biomedical semantics. 2016 Dec;7(1):1-9.

[6] Smith B, Ceusters W. Ontological realism: A methodology for coordinated evolution of scientific ontologies. Applied ontology. 2010;5(3–4):139.