

# NGBO: Introducing -omics data into biobanking ontology

Dalia Alghamdi<sup>1,3</sup>, Damion M. Dooley<sup>1</sup>, Gurinder Gosal<sup>1</sup>, Emma J. Griffiths<sup>1</sup>, William W.L. Hsiao<sup>1,2</sup>

<sup>1</sup>University of British Columbia, Vancouver BC V6T 1Z4, Canada.

<sup>2</sup>BC Centre for Disease Control Public Health Laboratory, Vancouver, BC V5Z 4R4, Canada.

<sup>3</sup>King Fahad Medical City, Riyadh 59046, Saudi Arabia.

## Abstract

*A biobank contains a collection of biological samples, along with associated medical information of sample donors, for further research use. Given the wealth of information that can be derived from stored information and biological materials, there is a pressing need for structuring biobank data in a manner that allows subsequent computer-based analyses. Currently, the value of biobank data lies in how it can be linked with other molecular and clinical data (“-omics data”), to provide new insights into health and disease. Linking data has thus far, however, proven challenging due to unstructured and incompatible data types. Here, we describe the development of a Next-Generation biobanking ontology (NGBO) that is capable of supporting both Biospecimen processing, management, storage and retrieval infrastructure, and acting as a knowledge hub for an integrated clinical and translational research ecosystem integrating -omics data.*

## Keywords:

Biobank, ontology, biomedical ontology.

## Introduction

A biobank consists of the various biological samples linked to the medical information of the sample donors to be used in translational and biomedical research. Possible sample components that can be stored in a Biobank include tissues, cells and fluids. The process of collecting human tissues to conduct biomedical research had begun centuries ago. However, medical research has undergone recent developments and technological advancement in data storage techniques, sampling and clinical testing. Today's biobanks involve the storage and management of highly complex data, including proteomics, genetic and -omics details among other data (De Souza and Greenspan 2013). The informatics needed in biobanks are more complex and advanced.

The primary purpose of a biobank is sample and data sharing. The value of biobank data could be optimized if the broad array of content of various biobank records could be standardized and integrated. A key development of facilitating data standardization includes the application of semantics and ontologies (Andrade et al. 2012). In this work, we integrated biobank -omics data through the creation of next-generation biobanking ontology in order to better structure data for various scientific research and personalized medicine analyses. NGBO harmonizes the instrumentation and procedures used to prepare and process specimens, and also covers terminology used to describe computational biology algorithms, analytical tools, electronic-communication protocols, in vitro assays along with

integrating Biorepository Accreditation Program (BAP) requirements.

## Methods

NGBO is being developed using the principles of Open Biological and Biomedical Ontologies (OBO) foundry. These principles include the application of the Basic Formal Ontology (BFO), a small upper level ontology mostly used to support information retrieval, analysis and integration across scientific and biomedical and other biological domains (Arp, Smith, and Spear 2015). Terms and logical relationships were re-used from Basic formal ontology (BFO), Ontology for Biobanking, Informed Consent Ontology (ICO), Ontology for Biomedical investigation (OBI) and Genomic Epidemiology Ontology (GENEPIO). Additional terms were identified and added to NGBO through the process of analyzing BAP requirement and use cases from King Fahad Medical City clinical genomics reports and standard operating procedures, mainly terms describing (-omics) data and processes, such as the description of the output data files of analytical bioinformatics process.

## Results

NGBO currently contains 1593 classes and 89 object properties. The ontology files are publicly available at <https://github.com/Dalalghamdi/NGBO>. NGBO depends primarily on the (is-a) relation between classes and subclasses, thereby providing a hierarchy of classes that also enables inheritance of the properties. Additionally, pre-existing relations such as (is\_specified\_input\_of) were used when appropriate. The ontology is still under development.

## Future work

NGBO is an ongoing project, and in the current stage we are enriching the ontology by mapping it to different use cases.

## References

1. Andrade, André Q., Markus Kreuzthaler, Janna Hastings, Maria Krestyaninova, and Stefan Schulz. 2012. “Requirements for Semantic Biobanks.” *Studies in Health Technology and Informatics* 180: 569–73.
2. De Souza, Yvonne G., and John S. Greenspan. 2013. “Biobanking Past, Present and Future: Responsibilities and Benefits.” *AIDS* 27 (3): 303–12.