



12th International Meeting on Microbial  
Epidemiological Markers (IMMEM XII)  
Dubrovnik, Croatia  
18–21 September 2019

216 Empowering data sharing for genomics-based public health surveillance using ontologies

Emma Griffiths<sup>1</sup>, Damion Dooley<sup>1</sup>, Gurinder Gosal<sup>1</sup>, Ivan Gill<sup>1</sup>, Shannon Russell<sup>2</sup>, Lauren Tindale<sup>2</sup>, Virginia Pichler<sup>2</sup>, Thomas Matthews<sup>3</sup>, Aaron Petkau<sup>4</sup>, Josh Adam<sup>4</sup>, Dan Fornika<sup>2</sup>, Geoff Winsor<sup>5</sup>, Finlay Maguire<sup>6</sup>, Brian Alcock<sup>7</sup>, Andrew McArthur<sup>7</sup>, Robert Beiko<sup>8</sup>, Morag Graham<sup>4</sup>, Fiona Brinkman<sup>5</sup>, Gary Van Domselaar<sup>4</sup>, William Hsiao<sup>1;2</sup>

<sup>1</sup>University of British Columbia, Pathology and Laboratory Medicine, Vancouver, Canada, <sup>2</sup>BC Centre for Disease Control, Vancouver, Canada, <sup>3</sup>Public Health Agency of Canada, National Microbiology Laboratory, Winnipeg, Canada, <sup>3</sup>Public Health Agency of Canada, National Microbiology Laboratory, Winnipeg, Canada, <sup>5</sup>Simon Fraser University, Molecular Biology and Biochemistry, Vancouver, Canada, <sup>6</sup>Dalhousie University, Faculty of Computer Science, Halifax, Canada, <sup>7</sup>McMaster University, Biochemistry and Biomedical Sciences, Hamilton, Canada, <sup>8</sup>Dalhousie University, Faculty of Computer Science, Hamilton, Canada

**Background:** The Integrated Rapid Infectious Disease Analysis (IRIDA) platform is a user-friendly, decentralized, open-source bioinformatics and analytical web platform to support multijurisdictional infectious disease outbreak investigations using genomics data (irida.ca). The ability to harmonize contextual data (sample metadata, lab, clinical and epidemiological data) across sectors and institutions is critical for interpreting genomics-based analyses used for surveillance and public health investigations. To harmonize and integrate contextual data, we have developed two ontologies called the Genomic Epidemiology Ontology (GenEpiO) and the Food Ontology (FoodOn). Ontologies provide hierarchies of well-defined, standardized vocabulary and built-in logic that enable comparisons of data at different levels of granularity, and enhance querying power.

Currently, we are developing ontology-driven tools and fine-grained access controls within IRIDA in order to enable different users to selectively share and access different fields of contextual information (Figure 1). Here, we describe four public health use cases that we have leveraged to expand our ontologies to i) better support One Health investigations of foodborne pathogens, and ii) better aggregate data for improved enterics and AMR prediction, iii) support access-control/tool development.

**Materials/methods:** Use cases focused on pathogens such as enterics, influenza and tuberculosis, and explore ontology-mediated contextual data exchange between Canadian agriculture and human health surveillance systems, as well as between other platforms and surveillance systems such as GenomeTrakr, Resistome Tracker, and others. User engagement, in the form of interviews and resource reviews, identified vocabulary gaps, as well as data exchange requirements. Stakeholder datasets were used to assess the utility of these ontologies for data harmonization and exchange.

**Results:** New vocabulary and logic identified during user engagement were integrated into our ontologies. LexMapr, an ontology-based data harmonization tool, was developed enabling the transformation of unstructured text to standardized terms, followed by mapping to 3rd party data specifications (<https://anaconda.org/bioconda/lexmapr>).

**Conclusions:** Our work indicated that ontology-based data harmonization enhances data interoperability and computability for different health systems. This work will be used to support access control development in IRIDA, which we anticipate will help create a trusted environment for enhanced cooperation and data sharing, which is crucial for moving towards global, genomics-based pathogen surveillance.

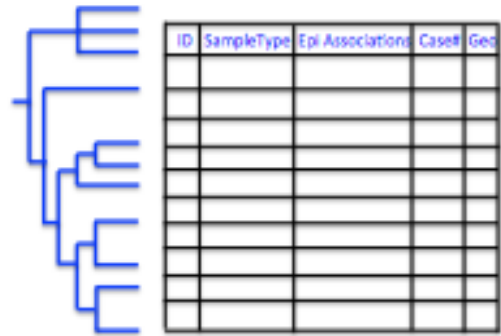
Figure 1

© ESCMID eLibrary by author

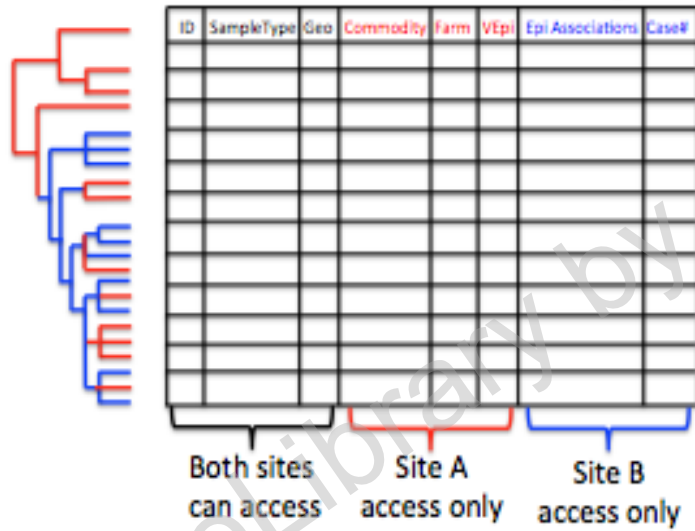
Instance A



Instance B



Shared Data Between Instances



© ESCMID eLibrary by author