Ontological Representation of CDC Active Bacterial Core Surveillance Case Reports

Albert Goldfain
Dept. of Eng. and Computer Science
Syracuse University
Syracuse, USA
agoldfai@syr.edu

Barry Smith
National Center for Ontological Research
Buffalo, USA
phismith@buffalo.edu

Lindsay G. Cowell
Dept. of Clinical Science
UT Southwestern Medical Center
Dallas, USA
lindsay.cowell@utsouthwestern.edu

I. INTRODUCTION

The Center for Disease Control and Prevention's Active Bacterial Core Surveillance (CDC ABCs) Program is a collaborative effort betweeen the CDC, state health departments, laboratories, and universities to track invasive bacterial pathogens of particular importance to public health [1]. The year-end surveillance reports produced by this program help to shape public policy and coordinate responses to emerging infectious diseases over time. The ABCs case report form (CRF) data represents an excellent opportunity for data reuse beyond the original surveillance purposes.

In this work, we focus on methicillin-resistant Staphylococcus aureus (MRSA), which has been tracked by the ABCs program since 2005. We use the Infectious Disease Ontology (IDO) Staphyloccocus aureus extension ontology (IDO-Staph), along with other ontologies following the principles of the Open Biomedical Ontologies Foundry (OBOF) to represent the entities referenced by the MRSA specific ABCs CRF. The goals of this effort are: (1) to demonstrate that infectious disease case report data can be positioned for reuse and linking to complementary data sources at the point of collection, (2) to identify any coverage gaps or limitations in the OBOF representation, and (3) to extend and reassess previous work in the ontology of infectious diseases [2,3,4].

II. ONTOLOGICAL REPRESENTATION OF CASE REPORTS

One of the unique problems for synthesizing surveillance data relating to any rapidly changing phenomenon of broad social impact – such as the rise of antibiotic resistance in bacteria – is that the range of information requires changes over time. In addition to temporal queries, such data frequently must be interrogated along several other dimensions: across pathogens in the ABCs program, across geographical regions represented by different ABCs surveillance sites and beyond, across pathogens with different forms of antibiotic resistance (with our evolving understanding of the mode of action and genetic basis for this resistance), and across data models / systems with different case definition criteria and different semantics for data entry fields.

The semantic web stack of technologies, when applied towards metadata representation and resource linking, is a particularly good fit for this task. The SPARQL query language for such representations also allows data to be stored in a decentralized manner. This is of particular importance for an international problem such as bacterial surveillance.

We use OBOF ontologies to represent: (1) entities referenced in the CDC ABCs CRF for MRSA, (2) the CDC case definition for MRSA infectious disease, and (3) the CDC inclusion criteria for cases.

IDO-Staph is an extension of the Infectious Disease Ontology (IDO) covering entities specific to *Staphylococcus aureus* infectious disease. Classes in IDO-Staph have supertypes in IDO, the Ontology for General Medical Science (OGMS), and BFO. Within this framework, many of the logical implications are inherited by descendant types from their supertypes. In creating our ontological representation to cover the ABCs CRF, we use the most specific OBOF term (i.e., the lowest descendent of a BFO term) that is applicable for each relevant entity in the CRF.

As an illustration of the sorts of entities to which the CRF needs to refer to, we represent the following information from a specific (hypothetical) case report:

John Doe is a 67 inch, 210 lb, 38 year-old patient at the Mayo Clinic with a case with a Staphylococcus aureus infectious disease. Labwork identified MRSA in a sample John's blood (MRSA bacteremia) after spa typing the isolate. The isolate was found to be SCCMec Type IV, tested positive for toxic-shock syndrome toxin, and negative for Panton-Valentin Leukocidn (PVL). This strain of isolate is known to be resistant to several antibiotics, including Methicillin. The underlying condition that led to the initial infectious disorder was intravenous drug use.

Our ontological representation of this case report is expressed as RDF triples. Relations are drawn from the OBO Relation Ontology [5]. The appropriate relationship between individuals referred to by CRF fields and the universals in OBOF ontologies is made explicit. For example, the following triples establish relationships between John Doe's particular disease and disorder, and the universal types they instantiate:

'John Doe's MRSA infectious disease' ***instance of***
    ido-staph:'staphylococcus aureus infectious disease'
'John Doe's MRSA infectious disorder' ***instance of***
    ido-staph:'staphylococcus aureus infectious disorder'
'John Doe's MRSA infectious disease' **has_material_basis**
    'John Doe's MRSA infectious disorder'.

The material basis of the infectious disease is the infectious disorder, which has as proper parts: an organism population of MRSA (i.e., the infection) and a portion of John Doe's blood. The MRSA isolate is sampled from a part of John Doe (his blood) and placed into culture. The time at which entities exist is very important here. The material sample isolated from John's disorder is no longer part of John (and thus no longer part of the material basis for his disease), but data derived from this sample can be predictive of the course of John Doe's disease, prognosis, and outcome. In culture, only a sub-population of MRSA organisms will have ever been a part of John, but we have faith in the stability of the predictions it allows because the salient properties of John's MRSA population are inherited by their immediate descendents. Thus, we can make inferences based on SCC*mec* and *spa* typing, toxin profiles, and other labwork assays performed on the isolate in culture, for example, the methicillin resistance of John's MRSA:

'John Doe's MRSA isolate' **has_disposition** 'John Doe's MRSA isolate's antibiotic resistance to methicillin'

'John Doe's MRSA isolate's antibiotic resistance to methicillin' ***instance of*** ido-staph:'PBP2a-mediated resistance to beta-lactam antibiotic'

IDO-Staph provides the ability to subtype specific drug resistance dispositions based on their mechanism of action. However much broader coverage is needed for the mechanisms of action involved for different antibiotics. The Comprehensive Antibiotic Resistance Database (CARD) and its associated ontology [6] provide a good start along these lines. To be brought fully into alignment with OBOF, CARD data would have to be linked to a suitable drug ontology such as DrON [7].

In virtue of its physical makeup, John Doe's MRSA infection (i.e., the population of MRSA organisms) has a *particular* antibiotic resistance towards methicillin. Moreover, resistance to methicillin can (and will) vary in degree across isolated samples. Laboratory personnel measure the degree of resistance by performing a minimal inhibitory concentration assay to produce a certain measurement datum.

We have elsewhere discussed the detailed representation of SCC*mec* types and toxin profiles for PVL and TSST in the context of the NARSA isolate repository, as well as the representational units for the lab processes and assays involved in classifying *Staphylococcus aureus* [4]. These representations are readily combined with the data from the CRF to enrich the clinical picture of John Doe's disease. The entities and relationships required for the ontological representation of SCC*mec* type IV (as in this case) are presented.

The infection type in this case is bacteremia, which is differentiated from other types of infection solely by anatomical location of the isolate (i.e., the bloodstream). The underlying condition listed for the MRSA infectious disease is intravenous drug use. Ontologically, this can be modeled as a disposition towards certain behaviors that would be explanatory for the how the MRSA came to be in John's bloodstream. Depending on the modeling needs, John's intravenous drug use can be associated with many other pieces of information (e.g., relating to the injection site bearing the portal of entry role for *Staphylococcus aureus*).

The final portion of the CRF is the classification of MRSA type. Absent any information that John Doe acquired MRSA while meeting the criteria for either HACO or HA, John's case would be classified as community-associated MRSA. Lists of criteria such as this are well suited for OWL/RDF since the task is determining if an instance satisfies a description.

III. A WEB-BASED MRSA CASE REPORTING SYSTEM

We have implemented a large part of the ABCs CRF for MRSA as a standards-compliant (HTML5/CSS3) web-based form.[1] The current version of the web-form is intended as a proof-of-concept for annotating CRF data at the point of collection. The web form is a custom solution rather than one built around a particular web framework. This allows for maximal flexibility in exporting to other data formats that are specifically required by external resources. The ultimate goal would be to implement such a system with direct EMR integration.

IV. CONCLUSION

Our annotation of such data with OBOF types and relations can provide several advantages, including: (1) precise semantics and definitions can be enforced during data entry, (2) linkage to other infectious disease resources, such as the CARD, to enable broader queries, (3) harmonization and comparability of multi-year CRF data (e.g., for a longitudinal study), (4) the possibility for retrospective application of new inclusion criteria, and (5) an OWL/RDF data model with which to build web applications around CRF data.

Most of the resources necessary to build an ontological representation of the entities referred to by the ABCs CRF are already part of ontologies conformant to OBOF principles. Some of the gaps in coverage include: (1) an ontological resource specifically for pathogen genes and gene products, (2) a drug ontology that classifies methods of action for different antibiotics, and (3) a good ontological relation template for how information about isolates in culture can lead to inferences about the disorders these bacteria are sampled from.

If case report data is properly represented and linked to other resources, this data can lead to insights beyond the original scope of CDC ABCs surveillance.

---

[1] See http://www.awqbi.com/ido/abccrf/

REFERENCES

[1] National Center for Immunization and Respiratory Diseases, Division of Bacterial Diseases, "CDC – ABCs: Overview – Background" , http://www.cdc.gov/abcs/overview/background.html , Retrieved Jan 22, 2014.

[2] A. Goldfain, B. Smith, and L. G. Cowell, "Towards an Ontological Representation of Resistance: The Case of MRSA", Journal of Biomedical Informatics, vol. 44(1), pp. 35–41, 2011.

[3] A. Goldfain, B. Smith, and L. G. Cowell, "Dispositions and the Infectious Disease Ontology", Proceedings of the Sixth International Conference on Formal Ontology in Information Systems, pp. 400–413, 2010.

[4] A. Goldfain, B. Smith, and L. G. Cowell, "Constructing a Lattice of Infectious Disease Ontologies from a Staphylococcus aureus Isolate Repository", Proceedings of the Third International Conference on Biomedical Ontology, 2012.

[5] B. Smith, W. Ceusters, B. Klagges, J. Köhler, A. Kumar, J. Lomax, C. Mungall, F. Neuhaus, A. L. Rector, and C. Rosse, "Relations in biomedical ontologies", Genome Biology, vol 6: R46.

[6] A. G. McArthur et al, "The Comprehensive Antibiotic Resistance Database", Antimicrobial Agents and Chemotherapy, vol. 57, pp. 3348–3357.

[7] W. R. Hogan, J. Hanna, E. Joseph, and M. Brochhausen, "Towards a Consistent and Scientifically Accurate Drug Ontology", Proceedings of the Fourth International Conference on Biomedical Ontology, 2013.

Ontological Representation of CDC Active Bacterial Core Surveillance Case Reports

Albert Goldfain[1], Barry Smith[2], Lindsay G. Cowell[3]

(1) Dept. of Eng and Computer Science, Syracuse University, (2) National Center for Ontological Research,
(3) Dept. of Clinical Science, UT Southwestern Medical Center

ABSTRACT

We propose an ontological representation to support the annotation of a CDC Active Bacterial Core surveillance (ABCs) case report form, specifically the form used for Methicillin-Resistant *Staphylococcus aureus* surveillance. The ontological representation is developed using source ontologies from the Open Biomedical Ontology Foundry. A prototype web-based case report form is implemented to demonstrate how the proposed ontology resource can support the automatic annotation of case report data. The prototype implementation can be found at http://www.awqbi.com/ido/abccrf/. We argue that the annotated data will enable reuse of the surveillance data beyond the original scope and purpose of its collection. Design considerations, benefits, and limitations of the ontological representation are described.

GOALS:
- Demonstrate that infectious disease case report data can be positioned for reuse and linking to complementary data sources *at the point of collection*
- Identify any coverage gaps or limitations in the OBOF representation, and
- Extend and reassess previous work in the ontology of infectious diseases.

CDC ABCs SURVEILLANCE PROGRAM

CDC Active Bacterial Core Surveillance (ABCs) program is a collaborative effort between the CDC, state health departments, laboratories, and universities to track invasive bacterial pathogens of particular importance to public health.

Case reports produced for six emergent pathogens: group A and group B Streptococcus, Haemophilus influenzae, Neisseria meningitis, Streptococcus pneumoniae, and methicillin-resistant Staphylococcus aureus (MRSA). The primary output of the ABCs program is a yearly epidemiological report on each of the pathogens covered.

Cases, Deaths & Inferred PFGE type by Epidemiological Classification

MRSA Class	No. (Rate) Cases[b]	No. (Rate) Deaths[c]	Inferred PFGE Type (%)			
			Tot N	USA100	USA300	USA500/ Iberian
CA	937 (4.8)	85 (0.4)	145	30 (20.7)	100 (69.0)	5 (3.5)
HCA[a]	3565 (18.1)	475 (2.4)	535	267 (49.9)	150 (28.0)	69 (12.9)
HO	788 (4.0)	120 (0.6)	110	52 (47.3)	38 (34.6)	11 (2.7)
HACO	2777 (14.1)	355 (1.8)	425	215 (50.6)	112 (26.4)	58 (13.7)

Excellent opportunity for case report data reuse.

In this work, we focus on ABCs MRSA case reports. The CDC lists some pathogen-specific objectives for MRSA surveillance in addition to the main ABCs program objectives:

1. To evaluate changes in rates of hospital-onset (HO), healthcare-associated community onset (HACO), and community-associated (CA) invasive [MRSA] disease over time and across different geographic areas
2. To identify populations at risk for invasive MRSA disease,
3. To describe the molecular and microbiologic characteristics of [HA], [HACO], and [CA] MRSA"

Achieving these goals also requires linking case report data to relevant molecular and microbiological information.

In addition to querying case report data across time, the data may need to be queried across several other dimensions:
- Across pathogens in the ABCs program
- Across geographical regions represented by different ABCs surveillance sites and beyond.
- Across pathogens with different forms of antibiotic resistance (with our evolving understanding of the mode of action and genetic basis for this resistance).
- Across data models / systems with different case definition criteria and different semantics for data entry fields.

IDO-STAPH AND THE OBO FOUNDRY

IDO-Staph is an extension of the Infectious Disease Ontology (IDO) covering entities specific to *Staphylococcus aureus* infectious disease. Classes in IDO-Staph have supertypes in IDO, the Ontology for General Medical Science (OGMS), and BFO. For example, the taxonomy leading to *Staphylococcus aureus* infectious disease is as follows:
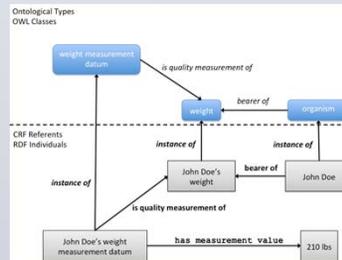
bfo:disposition
   ogms:disease
      ido:infectious disease
         ido-staph:staphylococcus aureus infectious disease

Within this framework, many of the logical implications are inherited by descendant types from their supertypes.

In creating our ontological representation to cover the ABCs CRF, we use the most specific OBOF term (i.e., the lowest descendent of a BFO term) that is applicable for each relevant entity in the CRF.

CASE REPORT REPRESENTATION

John Doe is a 67 inch, 210 lb, 38 year-old patient at the Mayo Clinic with a case with a Staphylococcus aureus infectious disease. Labwork identified MRSA in a sample John's blood (MRSA bacteremia) after spa typing the isolate. The isolate was found to be SCCMec Type IV, tested positive for toxic-shock syndrome toxin, and negative for Panton-Valentin Leukocidn (PVL). This strain of isolate is known to be resistant to several antibiotics, including Methicillin. The underlying condition that led to the initial infectious disorder was intravenous drug use.

Link to antibiotic resistance:
'John Doe's MRSA isolate' **has_disposition**
   'John Doe's MRSA isolate's antibiotic resistance to methicillin'
'John Doe's MRSA isolate's antibiotic resistance to methicillin' ***instance of***
   ido-staph:'PBP2a-mediated resistance to beta-lactam antibiotic'
'John Doe's MRSA isolate's antibiotic resistance to methicillin' **has_qualitative_basis** SOME
(is_quality_measured_as SOME
'methicillin minimal inhibitory concentration measurement datum of John Doe's MRSA isolate ')

'methicillin minimal inhibitory concentration measurement datum of John Doe's MRSA isolate'
instance of obi:'minimal inhibitory concentration'

Toxins and their dispositions
ido-staph:'Panton-Valentine leukocidin' *instance of* leukocidin
ido-staph:'Panton-Valentine leukocidin' *has_disposition* SOME ido:invasion disposition

pro:'toxic shock syndrome toxin-1' *instance of* protein
pro:'toxic shock syndrome toxin-1' *has_disposition* SOME ido:'exotoxin disposition'

PROOF OF CONCEPT IMPLEMENTATION

There are several immediate benefits of migrating the ABCs CRF from a paper form to an electronic web form. A web form would allow for form validation (on the client and server side), allow certain fields to be labeled as required input, and help to prevent data entry errors.

We have implemented a large part of the ABCs CRF (see http://www.awqbi.com/ido/abccrf/)
- Web form implementation
- Standards compliant (HTML5/CSS3)
- Follow-up questions as needed (jQuery)
- Client side logical constraints / required fields enforced
- RDF/XML output suitable for
  - Storage in a triplestore
  - SPARQL query
  - Input to a reasoner
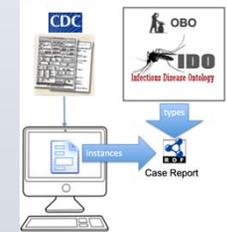
CONCLUSIONS

An ontological representation can also facilitate the extension, specialization, and linking of the CRF with different resources. Our annotation of such data with OBOF types and relations can provide several advantages, including:
- Precise semantics and definitions can be enforced during data entry.
- Linkage to other infectious disease resources, such as the Comprehensive Antibiotic Resistance Database, to enable broader queries.
- Harmonization and comparability of multi-year CRF data (e.g., for a longitudinal study).
- The possibility for retrospective application of new inclusion criteria. An OWL/RDF data model with which to build web applications around CRF data.

As we have seen, most of the resources necessary to build an ontological representation of the entities referred to by the ABCs CRF are already part of ontologies conformant to OBOF principles. Some of the gaps in coverage include:
1. An ontological resource specifically for pathogen genes and gene products,
2. A drug ontology that classifies methods of action for different antibiotics,
3. A good ontological relation template for how information about isolates in culture can lead to inferences about the disorders these bacteria are sampled from.

ACKNOWLEDGEMENTS AND CONTACT