Ontology representation and ANOVA analysis of vaccine protection investigation

Yongqun “Oliver” He

Unit for Laboratory Animal Medicine
Department of Microbiology and Immunology
Center for Computational Medicine and Bioinformatics
University of Michigan Medical School
Ann Arbor, MI 48109
I. Introduction of OBI, VO, and vaccine protection use case

II. Ontological representation of ANOVA

III. Ontological representation of *Brucella* vaccine protection investigation

IV. Ontology-based ANOVA analysis of *Brucella* vaccine protection investigation

V. Conclusion and Discussion
OBI: Ontology for Biomedical Investigations

• 19 communities trying to solve the same or related problems
• 5 year effort
• 2 phone calls per week, 2 meetings per year
• First stable release (Philly / 1.0) in Oct. 2009
• Release manuscript in revision
The Vaccine Ontology (VO)

• Aim: An ontology of the vaccine domain.
• Utilize the Basic Formal Ontology (BFO) as the top-level ontology
• Use OBI as another upper level ontology for vaccine investigation
• Follow OBO Foundry principles
• Multi-community and institutional collaboration

http://www.violinet.org/vaccineontology
VO Statistics

VO Terms

- **Class** (1266)
- **ObjectProperty** (12)

All Terms

- **Class** (1812)
- **ObjectProperty** (60)
- **DatatypeProperty** (4)

Imports

- imports: http://www.ifomis.org/bfo/1.1
  - Class (39)
- imports: http://purl.org/obo/owl/ro
  - ObjectProperty (24)
- imports: http://purl.cbolibrary.org/obo/iac/dev/iac-main.owl
  - Class (89)
  - ObjectProperty (8)
  - DatatypeProperty (4)
- imports: http://purl.cbolibrary.org/obo/vo/external/OBI_import.owl
  - Class (41)
  - ObjectProperty (12)
- imports: http://purl.cbolibrary.org/obo/vo/external/NCBITaxon_import.owl
  - Class (198)
- imports: http://purl.cbolibrary.org/obo/vo/external/PATO_import.owl
  - Class (17)
- imports: http://purl.cbolibrary.org/obo/vo/external/GO_import.owl
  - Class (2)
- imports: http://purl.cbolibrary.org/obo/vo/external/CHEBI_import.owl
  - Class (13)
- imports: http://purl.cbolibrary.org/obo/vo/external/DOID_import.owl
  - Class (57)
- imports: http://purl.cbolibrary.org/obo/vo/external/IDO_import.owl
  - Class (1)
- imports: http://purl.cbolibrary.org/obo/vo/external/ro_proposed_import.owl
  - ObjectProperty (9)
- imports: http://purl.cbolibrary.org/obo/vo/external/CARO_import.owl
  - Class (2)
- imports: http://purl.cbolibrary.org/obo/vo/external/FMA_import.owl
  - Class (2)
- imports: http://purl.cbolibrary.org/obo/iac/dev/ontology-metadata.owl
  - Class (5)

VO Browser: SPARQL supported VO visualization

http://www.violinet.org/vaccineontology/vobrowser
Vaccines Curated in VO

In total 301 vaccines (leaf nodes in VO)

- 151 licensed vaccines
- 23 host species
- Pathogens: viruses, bacteria, parasites
- Cancers: More coming
- Others: allergy, autoimmune disease, …
VO Imports >500 OBI and 13 Other Ontology Terms

• Import is better than reinvent existing terms. E.g.,

  + entity
    + occurrent
      + processual_entity
        + planned process
          + material processing
            + material combination
              + adding a material entity into a target
                + administering substance in vivo
                  + vaccination
                    - intraperitoneal vaccination

• Model ANOVA in OBI, then import to VO.
• How to import individual terms to VO? → OntoFox
• OntoFox: inspired by MIREOT & ontology modularization

Influenza Vaccine Protection Investigation

Three processes as described in the OBI JBMS paper:

- **vaccination**: a kind of *administering substance in vivo* process that realizes some *material to be added role*, borne by a *vaccine* (e.g., VacX) as well as a *target of material role* borne by an *organism* that also bears a *host role* (e.g., mouse).

- **pathogen challenge**: a kind of *administering substance in vivo* process. It realizes a number of roles - a *pathogen role* and *material to be added role* borne by the challenge *organism* (e.g., Influenza Virus), and a *target of material role* and *host role* borne by another organism (e.g., mouse).

- **survival assessment**: an *assay* that measures the *survival rate* (occurrence of death events) in one or more *organisms* that are monitored over time.

Vaccine protection investigation representation in JBMS OBI paper

Questions:

-- How to represent statistical analyses using ontology?

-- How to use ontology to represent instance data of biological investigations, e.g., vaccine protection investigation?

-- How to analyze instance data using ontology-based statistical analyses?
Outline

I. Introduction of OBI, VO, and vaccine protection use case

II. Ontological representation of ANOVA

III. Ontological representation of *Brucella* vaccine protection investigation

IV. Ontology-based ANOVA analysis of *Brucella* vaccine protection investigation

V. Conclusion and Discussion
Ontological Representation of Statistical Analyses

- OntoDM:
  - Ontological representation of data mining tasks and complex data types.
  - Align with OBI

- OBI statistical analysis:
  - Provide general top structure
  - Continuous efforts towards more details and deeper hierarchy
ANOVA: Analysis of Variance

• Aim: Test if the means of several groups are all equal.
• Includes statistical models, e.g., linear models.
• ANOVA runs F-test
• Data for ANOVA analysis:
  o Measurable data (e.g., time interval, vaccination dose)
  o Output of discretization of non-measurable data (e.g., mouse strains, gene mutants)
• ANOVA output: p-value
• Data sources:
  o Do experiments by ourselves
  o Extract data from journal articles
Ontology Representation of ANOVA

- ANOVA is a subclass of data transformation
- F-test is part of ANOVA
- ANOVA has specified input of data item
- The data item can be an output of a discretization process that discretizes non-measurable data
- To get data: data item extraction from journal article (IAO_0000443)
- ANOVA is concretization of ANOVA protocol
- ANOVA protocol includes a predictive model that specifies a testable hypothesis model
Ontology Design Pattern of ANOVA

- F-test
- ANOVA
  - has_part: concretization
  - has_specified_input: has_part
  - has_specified_output: p-value data item
- data item
  - has_specified_input: has_part
  - has_specified_output: discretization process
data transformation
- predictive model
  - has_part: categorical measurement datum
  - has_specified_output: data item
  - has_specified_input: has_part
- journal article
  - has_specified_input: has_part
  - has_specified_output: data item extraction from journal article
I. Introduction of OBI, VO, and vaccine protection use case

II. Ontological representation of ANOVA

√ III. Ontological representation of Brucella vaccine protection investigation

IV. Ontology-based ANOVA analysis of Brucella vaccine protection investigation

V. Conclusion and Discussion
Use case study: *Brucella* vaccine protection investigation

- *Brucella* spp. is intracellular, Gram negative bacteria that cause brucellosis, the most common zoonotic disease in the world, with 0.5 million new human cases yearly. It’s a even bigger problem for animals.

- No *Brucella* human vaccine available.
- *Brucella* cattle vaccines: RB51, strain 19, …
- Search PubMed “Brucella vaccine” → more than 300 publications.

- The efficacy of a *Brucella* vaccine candidate can be measured using a mouse model
Use case study: *Brucella* vaccine protection investigation

- *Brucella* does not kill mouse, so a mouse survival assay does not work.
- *Brucella* vaccine efficacy is measured by the reduction of colony forming units (CFU) of live *Brucella* in spleens of vaccinated mice compared to non-vaccinated mice.

- **Question:** What parameters (e.g., vaccination dose, mouse age) contribute to *Brucella* vaccine efficacy, and what not?
Q: What parameters are critical to RB51 vaccine efficacy?
**17 Tested Parameters**

<table>
<thead>
<tr>
<th>#</th>
<th>Classes / ANOVA variables</th>
<th>Sources &amp; term IDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>vaccine protection efficacy</td>
<td>VO: VO_0000456</td>
</tr>
<tr>
<td>2</td>
<td>vaccine strain</td>
<td>VO: VO_0001180</td>
</tr>
<tr>
<td>3</td>
<td>vaccine viability</td>
<td>VO: VO_0001139</td>
</tr>
<tr>
<td>4</td>
<td>vaccine protective antigen</td>
<td>VO: VO_0000457</td>
</tr>
<tr>
<td>5</td>
<td>mutated gene in vaccine strain</td>
<td>VO: VO_0001195</td>
</tr>
<tr>
<td>6</td>
<td>vaccination mouse strain</td>
<td>VO: VO_0001189</td>
</tr>
<tr>
<td>7</td>
<td>vaccination dose specification</td>
<td>VO: VO_0001160</td>
</tr>
<tr>
<td>8</td>
<td>pathogen strain for challenge</td>
<td>VO: VO_0001194</td>
</tr>
<tr>
<td>9</td>
<td>pathogen challenge (subclass)</td>
<td>OBI: OBI_0000712</td>
</tr>
<tr>
<td>10</td>
<td>CFU per volume</td>
<td>UO: UO_0000212</td>
</tr>
<tr>
<td>11</td>
<td>CFU reduction</td>
<td>VO: VO_0001164</td>
</tr>
<tr>
<td>12</td>
<td>IL-12 vaccine adjuvant</td>
<td>VO: VO_0001147</td>
</tr>
<tr>
<td>13</td>
<td>biological sex</td>
<td>PATO: PATO_0000047</td>
</tr>
<tr>
<td>14</td>
<td>vaccination (subclass)</td>
<td>VO: VO_0000002</td>
</tr>
<tr>
<td>15</td>
<td>animal age at vaccination</td>
<td>VO: VO_0000897</td>
</tr>
<tr>
<td>16</td>
<td>vaccination-challenge interval</td>
<td>VO: VO_0001191</td>
</tr>
<tr>
<td>17</td>
<td>challenge dose specification</td>
<td>VO: VO_0001161</td>
</tr>
</tbody>
</table>

**Q:** What parameters are critical to *Brucella* vaccine efficacy?
Outline

I. Introduction of OBI, VO, and vaccine protection use case

II. Ontological representation of ANOVA

III. Ontological representation of *Brucella* vaccine protection investigation

IV. Ontology-based ANOVA analysis of *Brucella* vaccine protection investigation

V. Conclusion and Discussion
Hypothesis

Some parameters in vaccine protection studies are critical in determining the result of *Brucella* vaccine protection efficacy in a mouse model, and some not.
Methods

Parameter selection and ontology representation

Data item extraction from journal articles

Data transformation: Discretization of non-measureable instance data

ANOVA analysis

Ontology representation of ANOVA output
### 17 Tested Parameters

<table>
<thead>
<tr>
<th>#</th>
<th>Classes / ANOVA variables</th>
<th>Sources &amp; term IDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>vaccine protection efficacy</td>
<td>VO: VO_0000456</td>
</tr>
<tr>
<td>2</td>
<td>vaccine strain</td>
<td>VO: VO_0001180</td>
</tr>
<tr>
<td>3</td>
<td>vaccine viability</td>
<td>VO: VO_0001139</td>
</tr>
<tr>
<td>4</td>
<td>vaccine protective antigen</td>
<td>VO: VO_0000457</td>
</tr>
<tr>
<td>5</td>
<td>mutated gene in vaccine strain</td>
<td>VO: VO_0001195</td>
</tr>
<tr>
<td>6</td>
<td>vaccination mouse strain</td>
<td>VO: VO_0001189</td>
</tr>
<tr>
<td>7</td>
<td>vaccination dose specification</td>
<td>VO: VO_0001160</td>
</tr>
<tr>
<td>8</td>
<td>pathogen strain for challenge</td>
<td>VO: VO_0001194</td>
</tr>
<tr>
<td>9</td>
<td>pathogen challenge (subclass)</td>
<td>OBI: OBI_0000712</td>
</tr>
<tr>
<td>10</td>
<td>CFU per volume</td>
<td>UO: UO_0000212</td>
</tr>
<tr>
<td>11</td>
<td>CFU reduction</td>
<td>VO: VO_0001164</td>
</tr>
<tr>
<td>12</td>
<td>IL-12 vaccine adjuvant</td>
<td>VO: VO_0001147</td>
</tr>
<tr>
<td>13</td>
<td>biological sex</td>
<td>PATO: PATO_0000047</td>
</tr>
<tr>
<td>14</td>
<td>vaccination (subclass)</td>
<td>VO: VO_0000002</td>
</tr>
<tr>
<td>15</td>
<td>animal age at vaccination</td>
<td>VO: VO_0000897</td>
</tr>
<tr>
<td>16</td>
<td>vaccination-challenge interval</td>
<td>VO: VO_0001191</td>
</tr>
<tr>
<td>17</td>
<td>challenge dose specification</td>
<td>VO: VO_0001161</td>
</tr>
</tbody>
</table>

**Dependent variable**

**Independent variables**
Instance Data

- 151 instance data were curated from 43 peer-reviewed papers.

- Each instance contains data for all 17 parameters.

- Non-measurable data (e.g., vaccine strain, vaccination route) are transformed to discretized data.
Instance Data in VO OWL file

Instance data in correct VO ontology hierarchy
Only related ontology terms are included

OntoBat: http://ontobat.hegroup.org/
ANOVA Analysis

• One-way ANOVA: a planned comparison
  – Compare “Significance level” with others
• R code for ANOVA analysis:
  ```r
  setwd(".");
  a<-read.table('vaxar.txt', header=T, sep="\t");
  library(dprep)
  b<-disc.ew(a,c(11,15,16,17,18))
  fit<-lm(CFU_dif ~ ., data=b)
  anova(fit)
  ```
• `disc.ew()`: discretization using equal width
• `lm()`: fit linear models
• “CFU_dif ~ .”: a formula: CFU_dif (response) vs other variables. This is the **predictive model**.
• `anova()`: Compute analysis of variance
Output Results for ANOVA Analysis

- ANOVA output results are presented by p-value and other values (e.g. F value).
- 6 parameters are not significant (P value > 0.05):
  - Mouse sex
  - IL-12 vaccine adjuvant
  - Vaccination route
  - Mouse age
  - Vaccination-challenge interval
  - Challenge dose
- 11 yes (P value < 0.05).

The ANOVA output results are also represented in ontology.
I. Introduction of OBI, VO, and vaccine protection use case

II. Ontological representation of ANOVA

III. Ontological representation of Brucella vaccine protection investigation

IV. Ontology-based ANOVA analysis of Brucella vaccine protection investigation

V. Conclusion and Discussion
Summary

• Ontological representation of ANOVA in OBI framework

• Ontological representation of *Brucella* vaccine protection investigation

• Ontology-based ANOVA analysis of *Brucella* vaccine protection investigations
Advantages of Ontology-based Statistical Analyses

- Allow data consistency checking
  - e.g., RB51 is a *Brucella* vaccine
  - BCG is a TB vaccine but not a *Brucella* vaccine
- Data sharing in Semantic Web
- Advanced data analysis in Semantic Web
- Automated reasoning
Future Work

• To replicate the statistical analysis, we will need to know which software and its version
  ➢ Software Ontology SWO may be used
• Represent the null hypothesis
• Represent different ANOVA
  – e.g., one-way, factorial ANOVA
  – e.g., linear model, randomization-based ANOVA
• Analyze more vaccine and other data
• Represent other statistical methods
Acknowledgements

Univ. of Michigan: Zuoshuang Xiang, Thomas Todd


The OBI Consortium, VO, IAO, PRIRN

NIH-NIAID Grant: R01AI081062