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

Outline

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

II. Ontological representation of ANOVA

III. Ontological representation of *Brucella* vaccine protection investigation

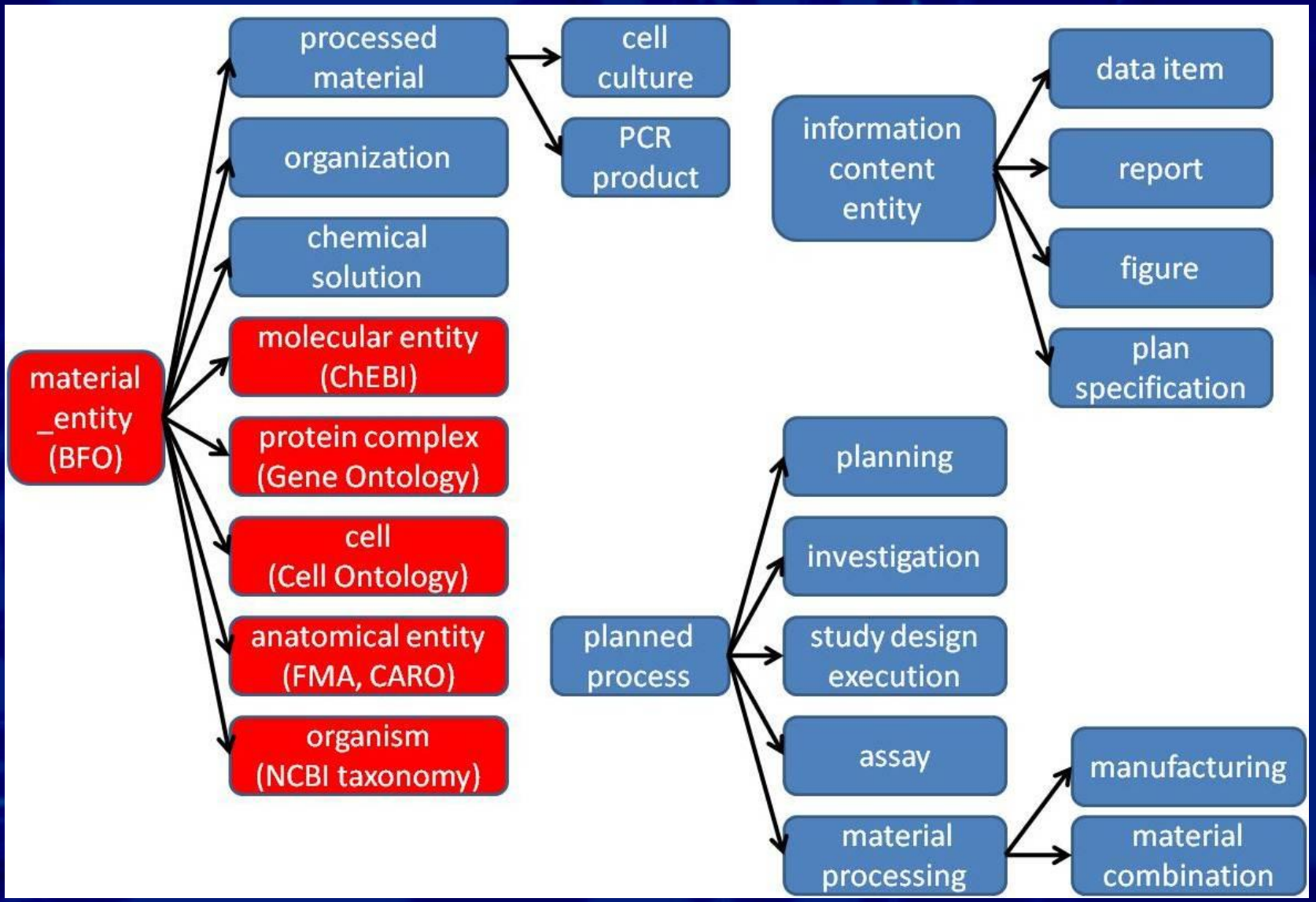
IV. Ontology-based ANOVA analysis of *Brucella* vaccine protection investigation

V. Conclusion and Discussion

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

OBI High Level Class Hierarchy



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

VO Browser: SPARQL supported VO visualization

http://www.violinet.org/vaccineontology/vobrowser

Vaccines Curated in VO

	Virus	Bacteria	Parasit	Cancer	Total
Human	70 (70)	73 (56)	5	1 (1)	138 (116)
Mouse	9	79	2	0	89
Guinea_pig	2	17	0	0	19
Cattle	1 (1)	8 (2)	5 (1)	0	13 (3)
Chicken	6 (1)	3	4	0	13 (1)
Canis	5 (3)	1 (1)	4	0	10 (4)
Pig	5 (5)	5 (3)	0	0	10 (8)
Bird	6 (5)	4 (4)	0	0	10 (9)
Monkey	6	2	2	0	10
Horse	5 (5)	2 (1)	1	0	8 (6)
Ovis	0	5	1	0	6
Rabbit	0	5 (1)	0	0	5 (1)
Fish	2 (1)	3 (2)	0	0	5 (3)
Rat	0	4	0	0	4
Felis	2 (1)	0	0	0	2 (1)
Chinchillas	0	1	0	0	1
Bubalus_bubalis	0	1	1	0	1
Ferret	1 (1)	0	0	0	1 (1)
Goat	0	1	0	0	1
Bison	0	1	1	0	1
Raccoon	1	0	0	0	1
Turkey	0	1	0	0	1
Deer	0	0	0	0	0
Total	108 (91)	181 (70)	23 (1)	1 (1)	301 (151)

-- In total 301 vaccines (leaf nodes in VO)

-- 151 licensed vaccines

-- 23 host species

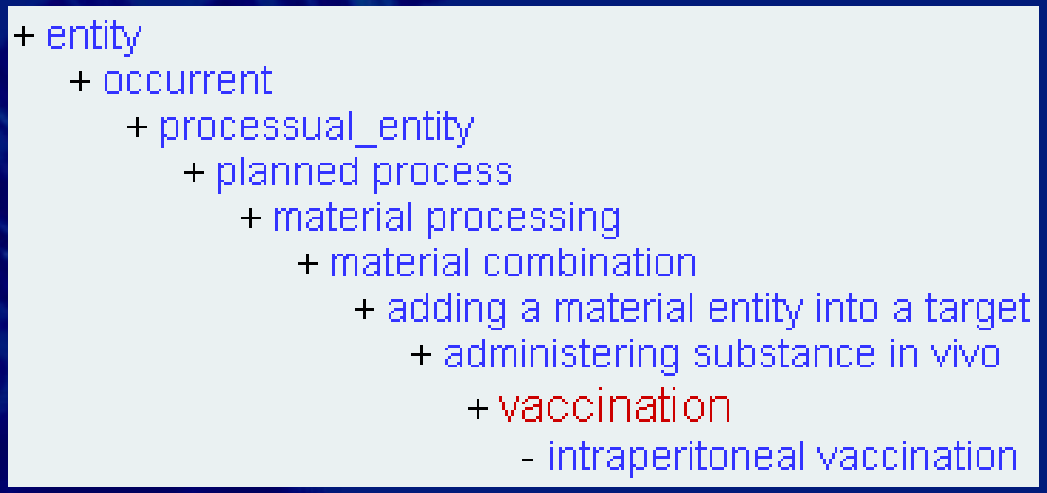
-- Pathogens: viruses, bacteria, parasites

-- 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.,



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

Reference: Xiang Z, Courtot M, Brinkman RR, Ruttenberg A, He Y. OntoFox: web-based support for ontology reuse. *BMC Research Notes*. 2010, **3:**175. [PMID: 20569493]

http://ontofox.hegroup.org/

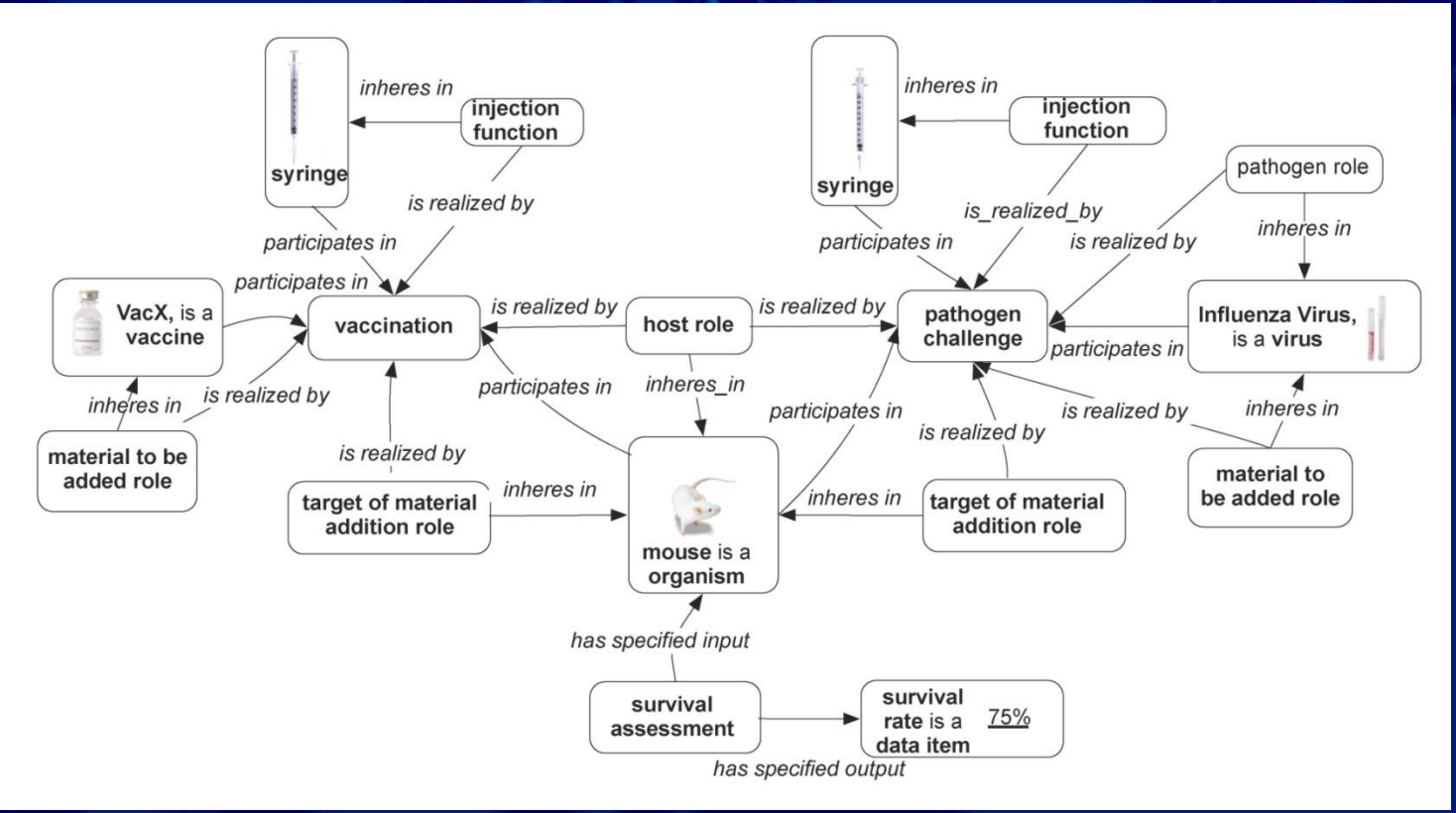
Influenza Vaccine Protection Investigation

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.

Reference: Brinkman RR, Courtot M, Derom D, Fostel JM, He Y, Lord P, Malone J, Parkinson H, Peters B, Rocca-Serra P, Ruttenberg A, Sansone S, Soldatova LN, Stoeckert Jr. CJ, Turner J, Zheng J, the OBI consortium. Modeling biomedical experimental processes with OBI. *Journal of Biomedical Semantics*. 2010, 1(Suppl 1):S7.

Vaccine protection investigation representation in JBMS OBI paper



Reference: Brinkman RR, Courtot M, Derom D, Fostel JM, He Y, Lord P, Malone J, Parkinson H, Peters B, Rocca-Serra P, Ruttenberg A, Sansone S, Soldatova LN, Stoeckert Jr. CJ, Turner J, Zheng J, the OBI consortium. Modeling biomedical experimental processes with OBI. *Journal of Biomedical Semantics*. 2010, 1(Suppl 1):S7.

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

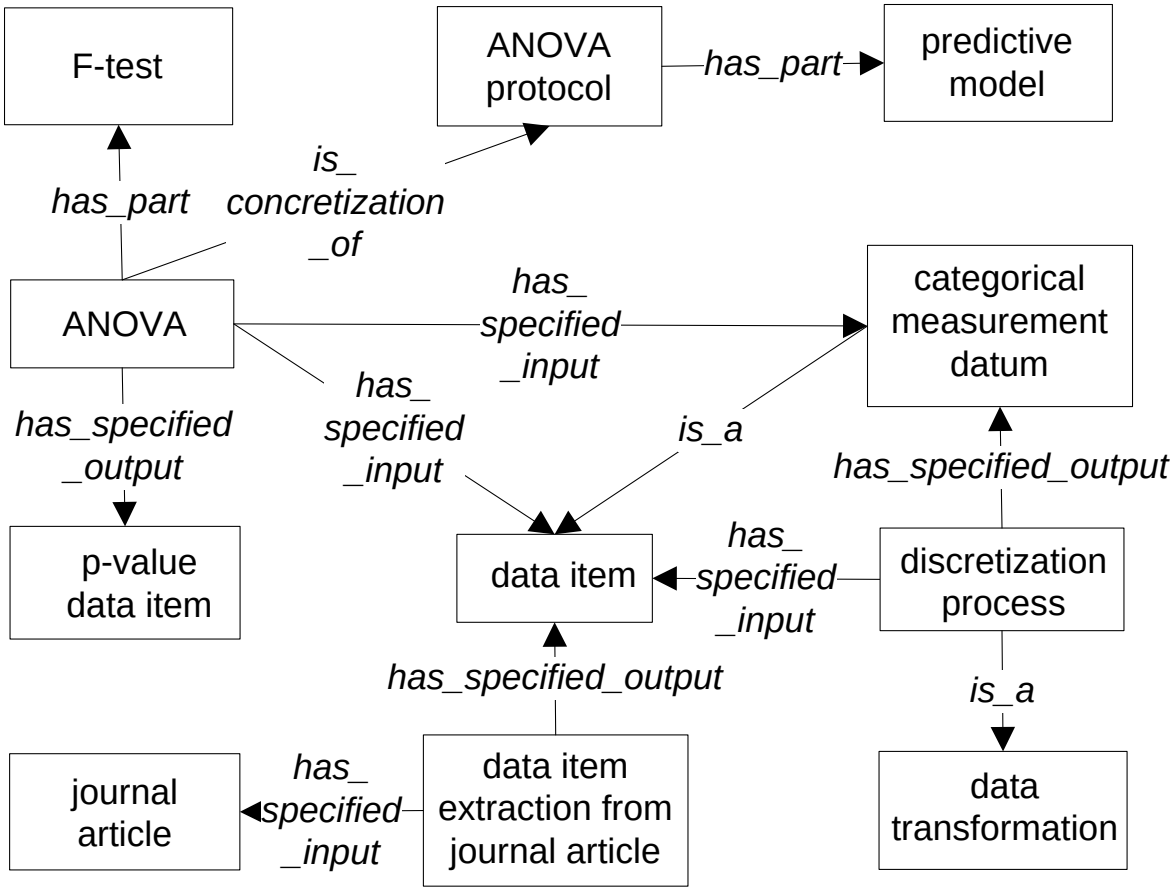
ANOVA: Analysis of Variance

- 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:
  - Measurable data (*e.g.*, time interval, vaccination dose)
  - Output of discretization of non-measurable data (*e.g.*, mouse strains, gene mutants)
- ANOVA output: p-value
- Data sources:
  - Do experiments by ourselves
  - 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 descretizes 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



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

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.

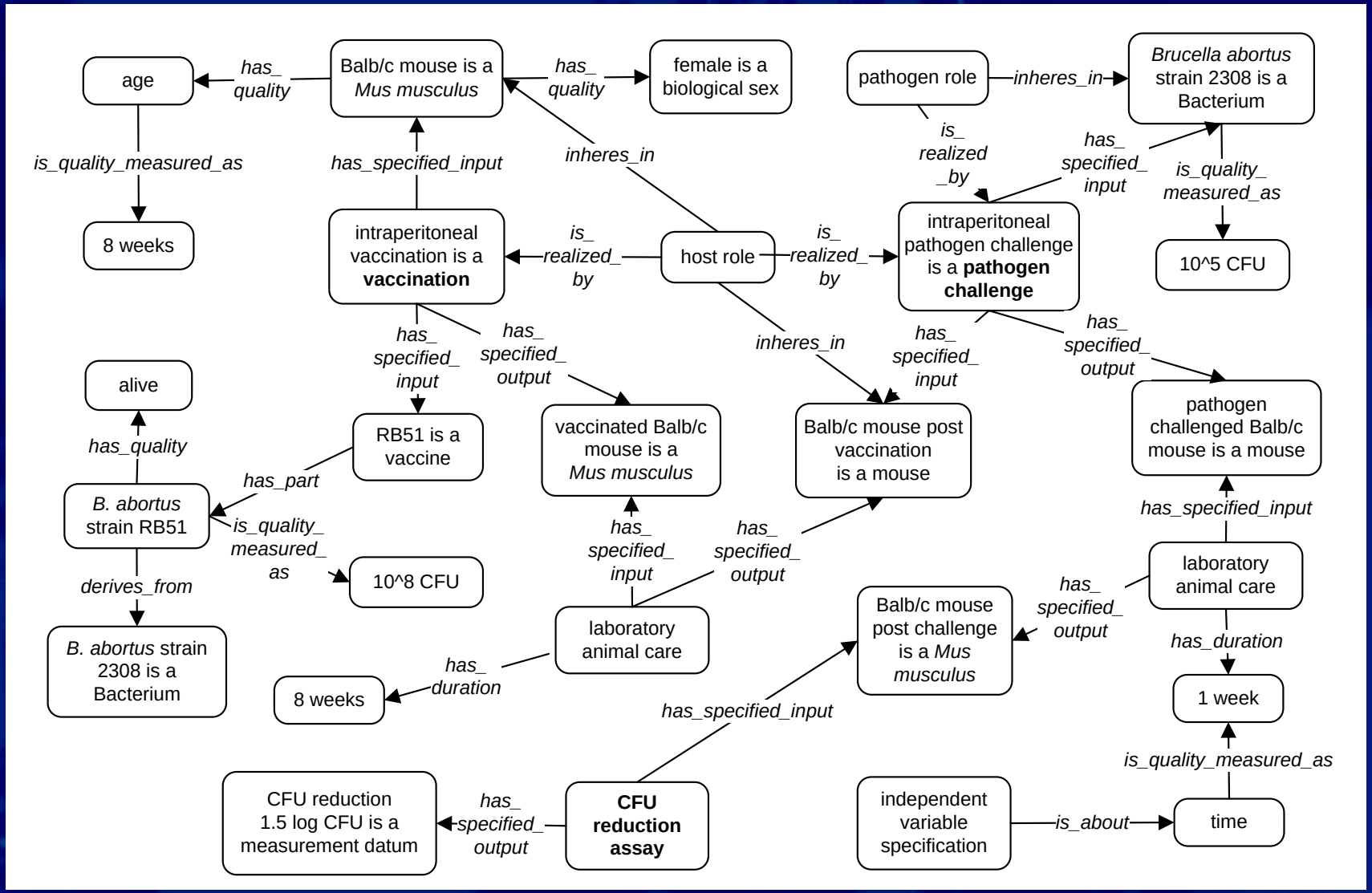
- The efficacy of a *Brucella* vaccine candidate can be measured using a mouse model

Use case study: *Brucella* vaccine protection investigation

- *Brucella* does not kill mouse, so a mouse survival assay do 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?

Example: *Brucella* Vaccine RB51



Q: What parameters are critical to RB51 vaccine efficacy?

17 Tested Parameters

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

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

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

Dependent variable

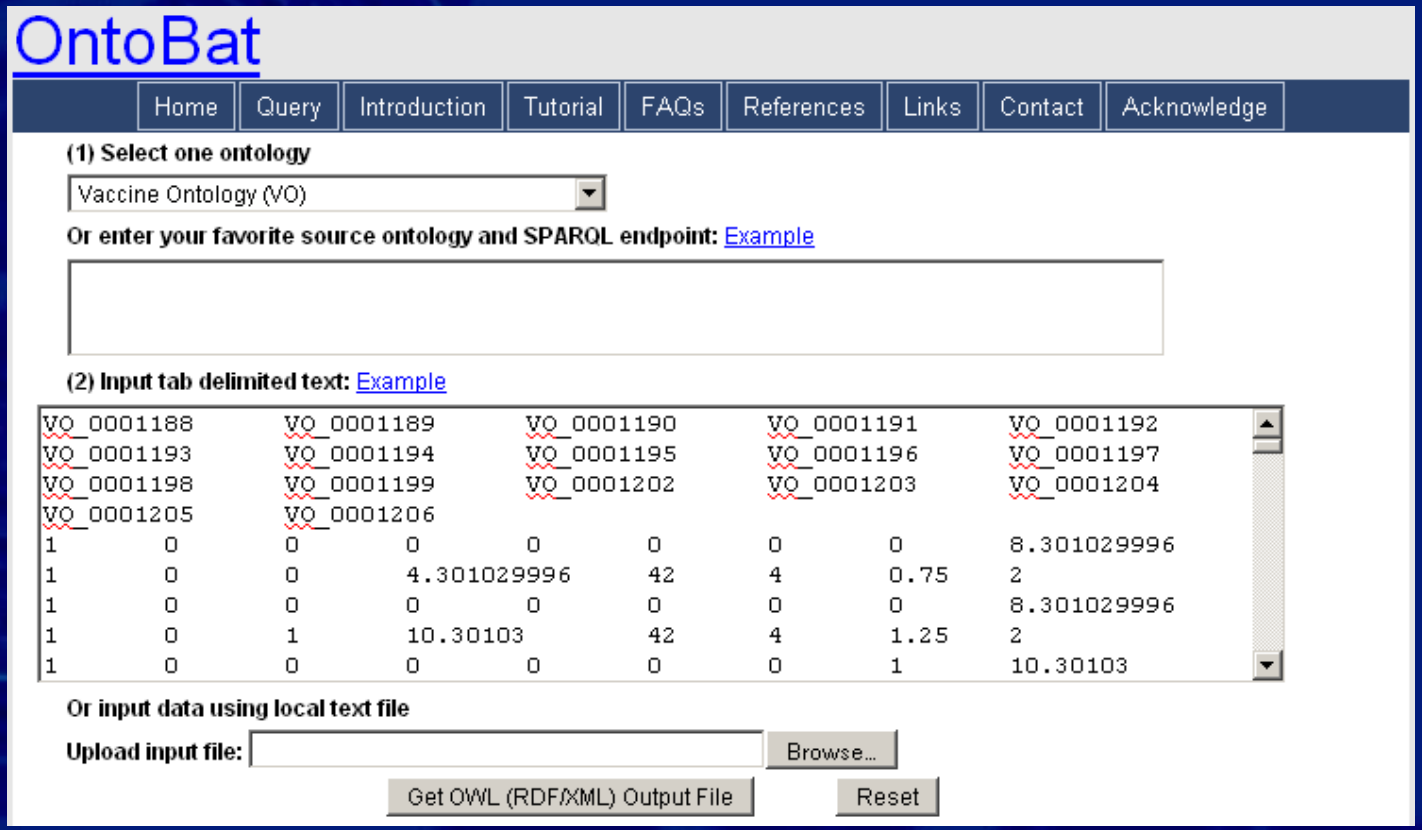
Independent variables

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 descretized 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:

```
1  setwd(".");
2  a<-read.table('vaxar.txt', header=T, sep="\t");
3  library(dprep)
4  b<-disc.ew(a,c(11,15,16,17,18))
5  fit<-lm(CFU_dif ~ ., data=b)
6  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

```
============================================================
Response: CFU_diff_sign
              Df  Sum Sq Mean Sq  F value    Pr(>F)
Vax_strain     1  9.4564  9.4564  83.0867 1.060e-15 ***
Live_dead      1  8.5377  8.5377  75.0152 1.374e-14 ***
IL.12          1  1.1211  1.1211   9.8504 0.0020917 **
SOD            1  4.6953  4.6953  41.2541 2.186e-09 ***
Gene_deleted   1  0.4697  0.4697   4.1265 0.0442085 *
Host_spp       1  0.0620  0.0620   0.5443 0.4619409
host_str       1  1.2265  1.2265  10.7765 0.0013135 **
sex            1  0.0028  0.0028   0.0244 0.8761106
Vax_route      1  0.0119  0.0119   0.1047 0.7468232
Vax_Dose_Log   1  0.5966  0.5966   5.2420 0.0236227 *
Vax_age_G      1  0.6202  0.6202   5.4491 0.0210768 *
Cha_strain     1  1.1622  1.1622  10.2116 0.0017433 **
Cha_route      1  0.1373  0.1373   1.2067 0.2739650
Cha_Dose_log   1  0.0738  0.0738   0.6488 0.4219775
Cha_interval   1  1.5008  1.5008  13.1866 0.0004013 ***
CFU            1 12.8479 12.8479 112.8854 < 2.2e-16 ***
CFU_dif        1  7.3974  7.3974  64.9956 3.831e-13 ***
Residuals    133 15.1372  0.1138
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- 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

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

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

OBI: Melanie Courtot, Ryan Brinkman, Jie Zheng, Christian J. Stoeckert Jr., James Malone, Philippe Rocca-Serra, Susanna-Assunta Sansone, Jennifer Fostel, Larisa N. Soldatova, Bjoern Peters, Alan Ruttenberg

The OBI Consortium, VO, IAO, PRIRN

NIH-NIAID Grant: R01AI081062