

Towards ontology enabled automated reasoning in environmental health

James A Overton
Jennifer Fostel
25 June, 2013

Summary of the Two-year Carcinogenicity Studies of beta-Picoline

	Male F344/N Rats	Female F344/N Rats	Male B6C3F1/N Mice	Female B6C3F1/N Mice
Level of carcinogenic activity	No evidence	Some Evidence	Equivocal Evidence	Clear Evidence

Recapping

- Counts of animal data in the different dose groups
 - Survival rates
- Comparison of treated to control group effects
 - Change in body weight relative to control group
- Identification of statistical trends
 - Dose-related trends of non-neoplastic and neoplastic effects
- Generalization of results to the identification of female mouse (B6C3F1/N) carcinogen

Raw materials

- SIFT (Simple Investigation Formatted Text)
- NTP data for individual animals in SIFT
 - Rat 1 (group1_cage1_1) exhibits mild lung hyperplasia
- Information about treatment groups in SIFT
 - Rat 1 was treated with vehicle only, and survived until the end of the study
- Statistical results for “dose-relatedness” of pathology incidence in SIFT
 - Lung hyperplasia is related to dose to a statistically significant extent
- Definition of NTP evidence levels

Start with SIFT Tabular Data (implicit relationships)

Participant characteristics (meta data)

\$GROUP_NAME	\$COMPARTOR_NAME	\$IS_CONTROL_GROUP	\$COMPOUND	\$DOSE	\$DOSE_UNIT	\$ROUTE	\$SPECIES_COMMON_NAME	\$SEX	\$STRAIN
GROUP8	GROUP2		beta-Picoline	1250	mg/l	DOSED WATER	Mouse	FEMALE	B6C3F1
GROUP7	GROUP1		beta-Picoline	1250	mg/l	DOSED WATER	Mouse	MALE	B6C3F1
GROUP6	GROUP2		beta-Picoline	625	mg/l	DOSED WATER	Mouse	FEMALE	B6C3F1
GROUP5	GROUP1		beta-Picoline	625	mg/l	DOSED WATER	Mouse	MALE	B6C3F1
GROUP4	GROUP2		beta-Picoline	312.5	mg/l	DOSED WATER	Mouse	FEMALE	B6C3F1
GROUP3	GROUP1		beta-Picoline	312.5	mg/l	DOSED WATER	Mouse	MALE	B6C3F1
GROUP2		T	beta-Picoline	0	mg/l	DOSED WATER	Mouse	FEMALE	B6C3F1
GROUP1		T	beta-Picoline	0	mg/l	DOSED WATER	Mouse	MALE	B6C3F1

PARTICIPANT_NAME	ORGAN	HISTOPATHOLOGY
GROUP1_CAGE1_1	Lung	Hyperplasia
GROUP1_CAGE1_1	Nose	Metaplasia
GROUP1_CAGE3_3	Adrenal cortex	Adenoma
GROUP1_CAGE3_3	Nose	Atrophy
GROUP1_CAGE7_7	Nose	Hyperplasia

Histopathology observations (data)

...convert to RDF triples with explicit relationships in the data

```
:group1-cage1-1
  rdfs:label "GROUP1_CAGE1_1" ;      # This subject
  rdf:type obo:NCBITaxon_10090 ;    # is a mouse
  sift:member-of :group1-cage1 ;    # member of sub-group :group1-cage1
  bfo:0000159 obo:PATO_0000384 ;    # has quality at all times: male
  sift:dose :dose-0 ;               # has dose :dose-0
  sift:death-status :terminal-sacrifice ; # was sacrificed
  sift:death-date :study-day-731 . # died on day 731

:group1-cage1-1-lung
  rdf:type obo:MA_0000415 ;          # This is a mouse lung
  sift:part-of :group1-cage1-1 .    # part of :group1-cage1-1

:group1-cage1-1-lung-hyperplasia
  rdf:type obo:MPATH_134 ;          # This is a mouse hyperplasia
  obo:BFO_0000052 :group1-cage1-1-lung ; # in :group1-cage1-lung
  sift:severity sift:mild .         # has severity: mild
```

Approach

- Build CEBS Application Ontology
 - Map terms in CEBS to external ontologies (OBO Foundry)
 - Build ontological definitions for internal CEBS terms
 - Align with OBI (Ontology for Biomedical Investigations)
- Build Template Engine

– Subject	SIFT	
This is a mouse	ncbi: 10090	rdf:type
named	PARTICIPANT_DISPLAY_NAME	rdfs:label
belongs to strain	STRAIN	cebs:strain
has quality at all times: sex	SEX	bfo:0000159
mamber of cage	CAGE	cebs:cage
member of group	GROUP	cebs:group
with dose	DOSE	cebs:dose
with dose unit	DOSE_UNIT	cebs:dose-unit
died on study day	DEATH_DATE	cebs:death-date
with death status	DEATH_STATUS	cebs:death-status
with chemical stressor	COMPOUND	(ChEBI or Pubchem)

Approach continued

- Build data representation for each SIFT file using the template rules

```
: {PARTICIPANT}
  rdfs:label "{$PARTICIPANT}" ;           # This subject
  rdf:type obo:NCBITaxon_10090 ;         # is a Mus musculus
  obo:BFO_0000086 {?$SEX} ;             # has quality {# $SEX}
  cebs: member-of : {$GROUP_NAME} ;     # is a member of { $GROUP_NAME}
  cebs:dose "{$DOSE} "^^xsd:float ;     # with group dose { $DOSE} mg/l
  cebs:death-status {?$DEATH_STATUS}    # died by {# $DEATH_STATUS}
```

- The template engine will use the rules to convert each SIFT file to an RDF representation (above).
- { } means we replace the content with the value from the SIFT file

Final aims

- RDF triple store
- This is Linked Data*, structured and computer-readable

* [Tim Berners-Lee](#) (2006-07-27). "[Linked Data—Design Issues](#)". [W3C](#). Retrieved 2010-12-18 at http://en.wikipedia.org/wiki/Linked_data

- Over which we can write SPARQL and other computational queries to produce summary reports
- Which will be available for use on the Semantic Web
- Which will be available for download
- Which users can query in new user interface

Quality Assurance

- Subject matter experts review mapping to pathology ontologies and disease ontologies
- CEBS Application Ontology will align with OBI, the Ontology for Biomedical Investigations,
 - OBI is a robust on-going community effort to create a common ontology for laboratory and clinical experimentation.
- Hire NTP to compare reasoned summary to NTP conclusion

Thanks!

- Vistronix contract HHSN273201000063U
 - Asif Rashid, Cari Favaro
 - James Overton (author)
- OBI, Ontology for Biomedical Investigations
 - <http://obi-ontology.org/>

