Semantic Assessment and Monitoring of Crowdsourced Geographic Information

Hamish McNair, Paul Goodhue
University of Canterbury
Christchurch, New Zealand
Outline

- Our research
- Project outline
- FOSS framework for the project
- Crowdsourcing information
- Determining Trust
- Ontologies
- Linked Data
- Future direction & Conclusion
Our Research

• Trusting Crowdsourced Geographic Information
  – Improving the trust of crowdsourced geographic information

• Crowdsourcing Spatial Data Supply Chains
  – Implications of trust beyond the capture of crowdsourced geographic information.
Project – Fruit Trees
Project – Fruit Trees
Project – Fruit Trees
Crowdsourcing

User Interface

WFS-T

Data Server

Database

OpenLayers

django

GeoServer

PostgreSQL
Data Model

```
<table>
<thead>
<tr>
<th>fruiting_observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>PK</td>
</tr>
<tr>
<td>fruiting_observation_id</td>
</tr>
<tr>
<td>fruit_tree_id</td>
</tr>
<tr>
<td>fruiting_observation</td>
</tr>
<tr>
<td>fruiting_observation_date</td>
</tr>
</tbody>
</table>

```

```
<table>
<thead>
<tr>
<th>fruit_tree</th>
</tr>
</thead>
<tbody>
<tr>
<td>PK</td>
</tr>
<tr>
<td>fruit_tree_id</td>
</tr>
<tr>
<td>fruit_tree_species</td>
</tr>
<tr>
<td>fruit_tree_height</td>
</tr>
<tr>
<td>fruit_tree_crown_diameter</td>
</tr>
<tr>
<td>fruit_tree_dbh</td>
</tr>
<tr>
<td>fruiting_observation</td>
</tr>
<tr>
<td>fruit_tree_trust_rating</td>
</tr>
<tr>
<td>geom</td>
</tr>
</tbody>
</table>

```

```
<table>
<thead>
<tr>
<th>fruit_tree_trust</th>
</tr>
</thead>
<tbody>
<tr>
<td>PK</td>
</tr>
<tr>
<td>fruit_tree_trust_id</td>
</tr>
<tr>
<td>fruit_tree_id</td>
</tr>
<tr>
<td>fruit_tree_trust_metrics_rating</td>
</tr>
<tr>
<td>fruit_tree_trust_fruiting_rating</td>
</tr>
<tr>
<td>fruit_tree_trust_location_rating</td>
</tr>
<tr>
<td>fruit_tree_trust_rating</td>
</tr>
<tr>
<td>fruit_tree_trust_rating_date</td>
</tr>
</tbody>
</table>
```
Conceptual Trust Model

Intrinsic:
- Spatial: Shape metrics of the geometry based on geometry type
- Temporal: Assessment of feature changelog or age of feature

Extrinsic:
- Spatial: Spatial comparison to neighbours based on rules about the CGI
- Temporal: Temporal comparison to neighbours based on rules about the CGI

Components of CGI:
- Spatio-temporal:
- Semantic:
- Social:

Assessments of the Information Source:
- Assessment of the author’s trust and likely influence on the trust of the CGI, e.g. through previous trust ratings or assessments of local knowledge
- Assessment of CGI to external data and ontologies known to influence the CGI
- Assessment of the trust of the author as reviewed by the crowd, e.g. through Linus’ Law, peer reviews and Consensus Crowdsorucing
Trust Model

PostgreSQL/PostGIS

Features queried From PostgreSQL

Python

Feature type rules queried from OWL

OWL

Comparisons between Features and ontology in python

Trust rating written Back to database
Feature Trust Rating

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>fruit_tree_species</td>
<td>Lemon</td>
</tr>
<tr>
<td>fruit_tree_height</td>
<td>2m</td>
</tr>
<tr>
<td>fruit_tree_crown_diameter</td>
<td>1m</td>
</tr>
<tr>
<td>fruit_tree_dbh</td>
<td>0.12m</td>
</tr>
<tr>
<td>fruiting_observation</td>
<td>Fruiting</td>
</tr>
<tr>
<td>fruit_tree_trust_rating_overall</td>
<td>100</td>
</tr>
<tr>
<td>fruit_tree_trust_rating_metrics</td>
<td>100</td>
</tr>
<tr>
<td>fruit_tree_trust_rating_fruiting</td>
<td>100</td>
</tr>
<tr>
<td>fruit_tree_trust_rating_location</td>
<td>100</td>
</tr>
</tbody>
</table>
## Feature Trust Rating

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>fruit_tree_species</td>
<td>Coconut</td>
</tr>
<tr>
<td>fruit_tree_height</td>
<td>5m</td>
</tr>
<tr>
<td>fruit_tree_crown_diameter</td>
<td>2m</td>
</tr>
<tr>
<td>fruit_tree_dbh</td>
<td>0.3m</td>
</tr>
<tr>
<td>fruiting_observation</td>
<td>Fruiting</td>
</tr>
<tr>
<td>fruit_tree_trust_rating_overall</td>
<td>66.67</td>
</tr>
<tr>
<td>fruit_tree_trust_rating_metrics</td>
<td>100</td>
</tr>
<tr>
<td>fruit_tree_trust_rating_fruiting</td>
<td>100</td>
</tr>
<tr>
<td>fruit_tree_trust_rating_location</td>
<td>0</td>
</tr>
</tbody>
</table>
Ontologies

• Ontologies in crowdsourcing?
  – accessibility
  – adjustability
  – versatility

• Implementation
  – Protégé
  – OWL/RDFS/XML
Ontology
Ontology

hasMaxHeight
Ontology

hasMaxHeight = 10 metres
Protégé

Class hierarchy:
- Thing
- measurement
- species
- tree
  - apple_tree
  - apricot_tree
  - coconut_tree
  - lemon_tree

Annotations: apple_tree
- label
  - apple_tree

Description: apple_tree
- Equivalent To
  - SubClass Of
    - hasFruitEnd value AppleFruitEnd
    - hasFruitStart value AppleFruitStart
    - hasMaxCrownSpread value AppleCrownSpread
    - hasMaxDBH value AppleDBH
    - hasMaxHeight value AppleHeight
    - hasMaxLat value AppleLatMax
    - hasMinLat value AppleLatMin
    - hasSpecies only apple_species
    - tree

General class axioms
Protégé
SPARQL Query in RDFLib

• Return reference attributes (via URIs)

```sparql
SELECT ?O
WHERE {
  <http://somethingGoesHere.org/foss4tree#appleTree> foss4tree:hasMaxHeight ?O
}
```
SPARQL Query in RDFLib

• Return reference attributes (via URIs)

```
SELECT ?O
WHERE {

<http://somethingGoesHere.org/foss4tree#appleTree> foss4tree:hasMaxHeight ?O

?O
}
```
SPARQL Query in RDFLib

• Return reference attributes (via URIs)

SELECT ?O
WHERE {

<http://somethingGoesHere.org/foss4tree#appleTree>

foss4tree:hasMaxHeight

?O

}
SPARQL Query in RDFLib

• Return reference attributes (via URIs)

```
SELECT ?O
WHERE {

<http://somethingGoesHere.org/foss4tree#appleTree> foss4tree:hasMaxHeight ?O
}
```
SPARQL Query in RDFLib

• Return reference attributes (via URIs)

```sparql
SELECT ?O
WHERE {
  <http://somethingGoesHere.org/foss4tree#appleTree> foss4tree:hasMaxHeight ?O
}
```
SPARQL Query in RDFLib

• Return reference attributes (via URIs)

```
SELECT ?O
WHERE {

<http://somethingGoesHere.org/foss4tree#appleTree>

foss4tree:hasMaxHeight

?O

}
```

TO THE TRUST MODEL
Linked Data

• Structure of RDF
  – Triples (Subject, Predicate, Object)

  <http://somethingGoesHere.org/foss4tree#t44>
  <foss4tree:hasHeight>
    <2.5>

  – Familiar (URIs), accessible, mashups
PYTHON MODEL

TRUST RATING > 70

WINDSPEED

MAP THIS

OUTPUT
?id <http://somethingGoesHere.org/foss4tree#hasTR> ?tr .
FILTER (?tr > 70)

?id <http://somethingGoesHere.org/foss4tree#hasSpecies> ?species .
?id <http://somethingGoesHere.org/foss4tree#hasFruiting> ?fruiting .
?id <http://somethingGoesHere.org/foss4tree#hasLat> ?lat .
?id <http://somethingGoesHere.org/foss4tree#hasHeight> ?height
LINKED DATA

WUNDERGROUND

PYTHON MODEL

TRUST RATING > 70

ID i    LAT i    LONG i    …
ID ii   LAT ii   LONG ii   …
ID iii  LAT iii  LONG iii  …
ID iv   LAT iv   LONG iv   …

OUTPUT
WEATHER UNDERGROUND

http://api.wunderground.com/api/##/geolookup/q/%f,%f.json

http://api.wunderground.com/api/##/conditions/q/pws:%s.json

www.wunderground.com

FOLIUM

OUTPUT
LINKED DATA

PYTHON MODEL

TRUST RATING > 70

WINDSPEED

WUNDERGROUND

OUTPUT
map1 = folium.Map(location = [Lat,Long], zoom_start=16)

For tree in trees:
    map1.simple_marker(treeLat, treeLon, popup = '...')

https://github.com/python-visualization/folium
LINKED DATA → PYTHON MODEL → WUNDERGROUND → OUTPUT

html ...
Lemon Tree (3 metres tall): In Fruit? - Yes! Windspeed: 4.8 km/hr
Where to from here...
WHERE TO FROM HERE...

WHY?

*Improved credibility of crowdsourced data*
Where to from here...

WHY?

*Improved credibility of crowdsourced data*

HOW?

*Trust models and implementation*
Where to from here...

WHY?
*Improved credibility of crowdsourced data*

HOW?
*Trust models and implementation*

THE HERE AND NOW
Traditional Spatial Datasets

- Credibility from legacy
- Provenance for tracing errors
- Dataset-level consideration
W3C PROV

DATASET

wasGeneratedBy

COLLECTION
W3C PROV

... back to triples!
Authoritative Data

• Dataset-level reactive provenance
Authoritative Data

- Dataset-level reactive provenance
Authoritative Data

- Dataset-level reactive provenance
Crowdsourced Data

- Feature level
Crowdsourced Data

• Feature level
Crowdsourced Data

- Feature level
Crowdsourced Data

- Feature level
Crowdsourced Data

- Feature level
Trust Ratings

• Simple indication of credibility of Datasets
  Features
  Attributes

• Provides proactive provenance

• Increases usability of crowdsourced data