

Social Web Meets Sensor Web: From User-Generated Content to Linked Crowdsourced Observation Data

Dong-Po Deng^{1,3}, Guan-Shuo Mai², Tyng-Ruey Chuang¹,
Rob Lemmens³, Kwang-Tsao Shao²

¹Institute of Information Science and ²Biodiversity Research Center,
Academia Sinica, Taiwan

³Faculty of Geo-Information Science and Earth Observation (ITC),
University of Twente, Netherlands



Background & Motivation

Citizen Science

- Citizen science refers to a distributed, collaborative problem-solving model in which a crowd of undefined size is engaged to solve a complex or scientific problem through an open call
 - Projects can happen at any scale—local, national, global
 - Projects generally take place over the long term tracking trends/changes over time

Galaxy Zoo

CLASSIFY STORY SCIENCE **GALAXY ZOO** DISCUSS PROFILE LANGUAGE

Few have witnessed what you're about to see

Experience a privileged glimpse of the distant universe as observed by the SDSS, the Hubble Space Telescope, and UKIRT

We are trying something new! Come help us understand a very specific type of galaxy and experience science from start to end. [Take part](#)

Classify Galaxies

To understand how galaxies formed we need your help to classify them according to their shapes. If you're quick, you may even be the first person to see the galaxies you're asked to classify.

[Begin Classifying](#)



How Do Galaxies Form?

Roughly one hundred billion galaxies are scattered throughout our observable Universe, each a glorious system that might contain billions of stars. Many are remarkably

History of Galaxy Zoo

The launch of this new version of Galaxy Zoo, the 4th, comes just a few weeks after the site's 5th birthday. It all started back in July 2007, with a data set made up of a million

<http://www.galaxyzoo.org/>

Audubon

Audubon Magazine | Audubon en Español | Newsroom | Audubon Near You | Chapter Services | Member Center | Careers | Contact Us

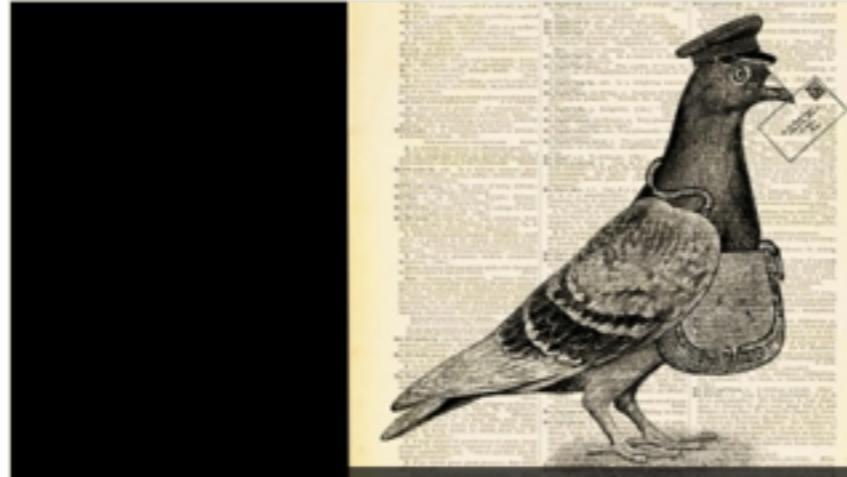
 **Audubon**

FOLLOW US:     



SEARCH

BIRDS CONSERVATION EDUCATION GET OUTSIDE ABOUT US DONATE TAKE ACTION


We're Reinventing Audubon Online. Any Suggestions?
We are transforming our website this summer and we want your thoughts, ideas, questions, and concerns along the way. What would make Audubon.org worth your time? How can it serve you? Bring it. [Read More](#)

MadameBricolage.net


Heather Paul/Flickr/CC
Support Audubon's efforts in Galveston and wherever birds are in danger.

[DONATE](#)


Hummingbirds at Home


Galveston Bay Oil Spill


Oil Spill Map


New Audubon Website


Photo of the Day

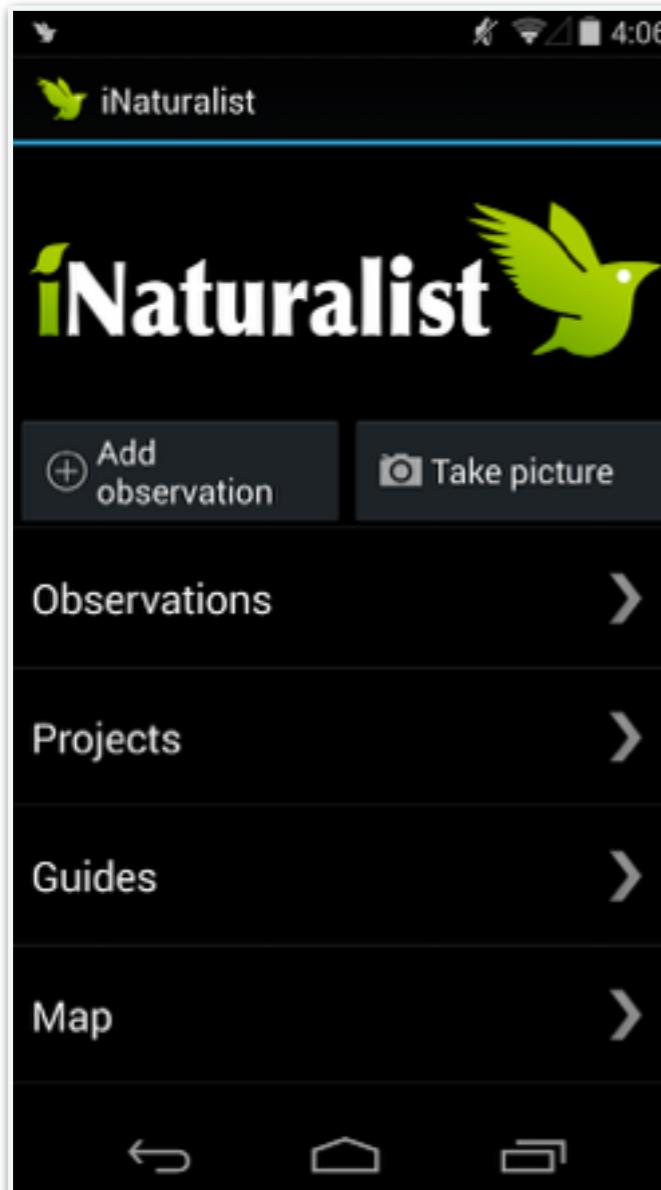

Audubon newsletter


Have you adopted a bird today?

[Audubon and Conservation News](#)

<http://www.audubon.org/>

iNaturalist



A screenshot of the iNaturalist.org website showing a project page. The header includes the iNaturalist logo, a search bar, and links for "Observations", "Species", "Projects", "Places", "Guides", "People", "Sign In or Sign up", and "Join this project". The main content area features a title "ASC Roadkill Observations" with logos for "ADVENTURERS & SCIENTISTS FOR CONSERVATION" and "Road Ecology Center". It includes a map of North America with several blue location pins. A callout says "Click here to visit the project on www.adventureandscience.org". To the right, there's a sidebar with links for "Observations / Map" (1465 observations), "Checklist" (287 taxa observed), "Journal", "Top contributors" (213 members), and a "View leaderboard" link. The "Top contributors" section lists three users: 1. toby (73 species, 159 observations), 2. greglasley (51 species, 193 observations), and 3. danjleavitt (38 species, 65 observations). At the bottom, there's an "About" section with a thank you message for roadkill observations.

<http://www.inaturalist.org/projects/asc-roadkill-observations>

Reptile Road Mortality

<https://www.facebook.com/groups/roadkilled/>

四處爬爬走(路殺社, Reptile Road Mortality)

Dongpo Deng Edit Profile

News Feed Messages Events 3

GROUPS

- OpenStreetMap台灣 2
- OpenData / Taiwan
- 四處爬爬走(路殺社, R...)
- 青平台／行動者 7
- 臺灣防災產業協會 20+
- 中研院羽球社 (Acade...
- 路況社 4
- OpenStreetMap US 5
- Open Data 颱風情報...
- State of the Map co... 3
- 明道中學 '91 四班
- 草屯65幫
- 特生中心資料管理與...
- SotM Asia Organizing...
- Semantic eScience
- Create Group

APPS

- Games 20+
- Twitter
- Games Feed 20+

FRIENDS

- Enschede, Neth... 20+

台灣野生動物路死觀察網 Taiwan Roadkill Observation Network
<http://roadkill.tw>

Arthur Wang CC BY SA 3.0 王嶽

Write Post Add Photo / Video Ask Question Add File

Write something...

陳秀晴 3 hrs · Edited

2014年4月5日, 下午 06:05:56
23.563038,120.414275



ABOUT 4,814 members

Open Group

社團四大宗旨：
1.改善路死 – 藉由大量的資料蒐集，分析路死嚴重的路段、季節及種類，改善道路設計或增加廊道、圍籬等設施，以減少野生動物因國內道路開發與車流量日漸增加而造成的直接死亡 (roadk... See More)

4,814 members (402 new) · Invite by Email

+ Add People to Group

SUGGESTED GROUPS See All

- Activity & Job (生態保育、環境教育、動物保護)
Dali Lin and 4 other friends joined + Join
- 社團法人台灣環境資訊協會(TEIA)
Rex Tsai and 32 other friends joined + Join

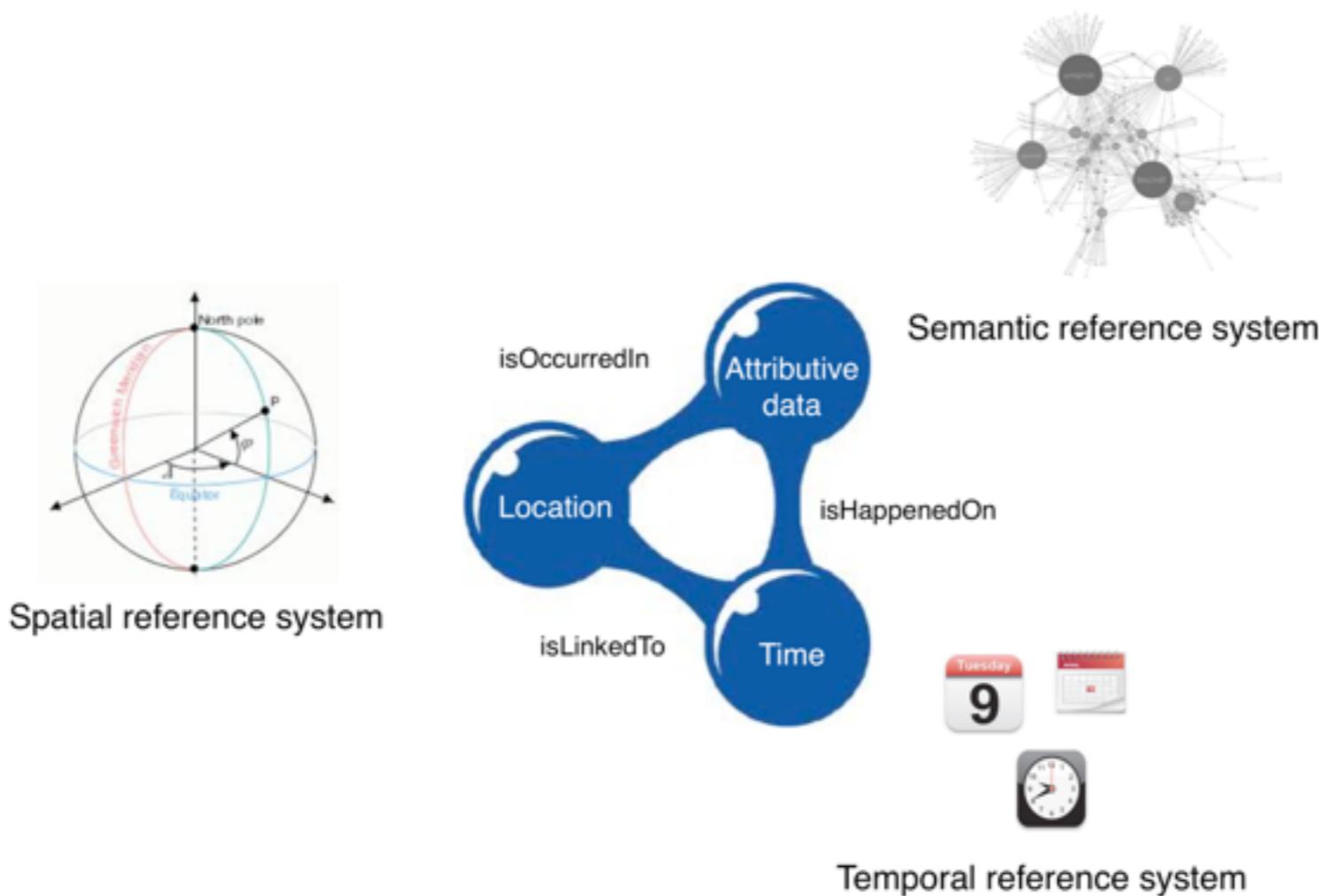
<https://www.facebook.com/groups/roadkilled/>

Motivation

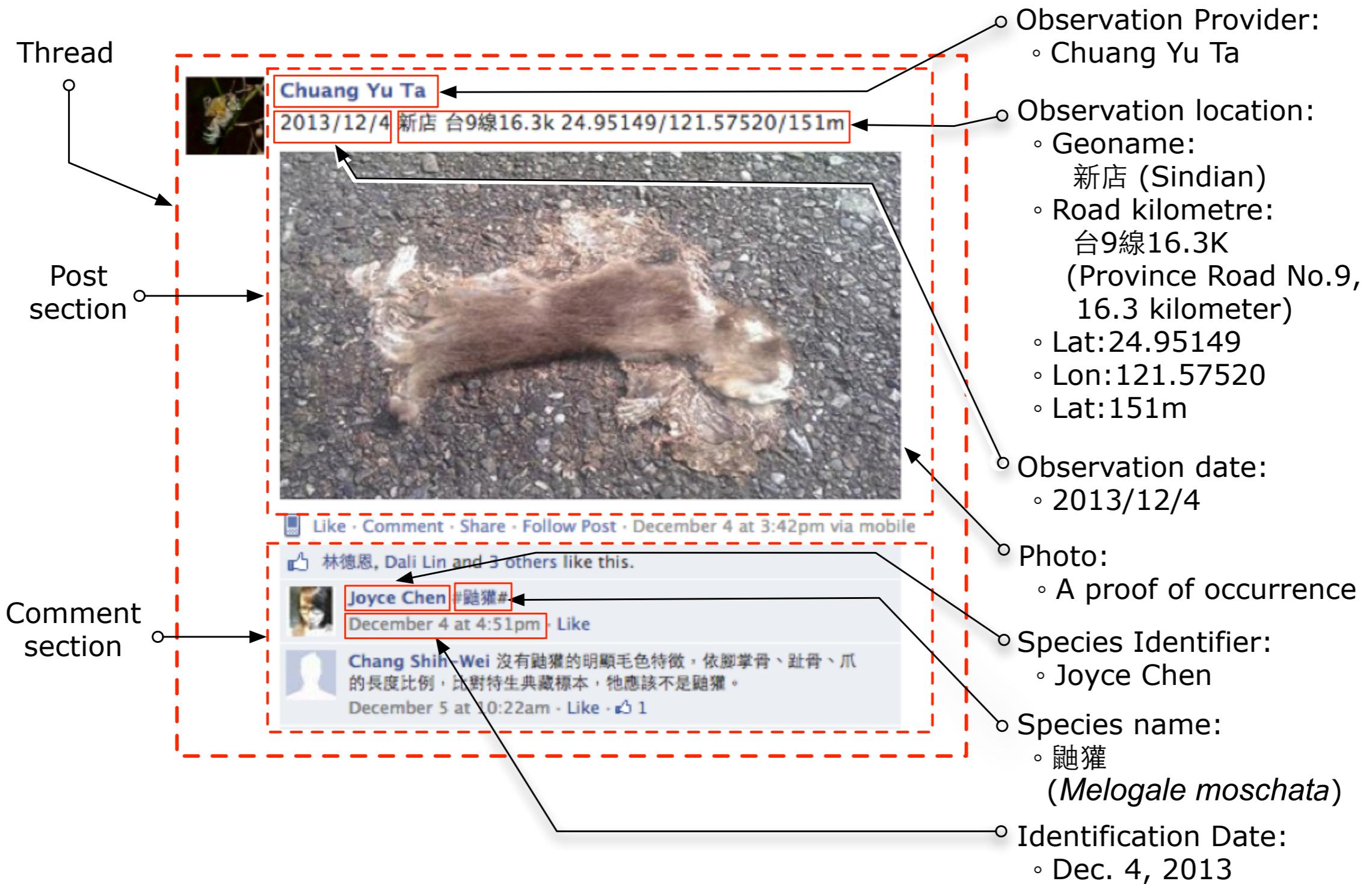
- By utilizing social media, citizen science projects can potentially engage many participants to contribute their observations covering a large geographic region and over a long time period.
- This is an improvement, for example, over traditional biodiversity surveys which typically involve relatively few people in confined regions and periods

Motivation

- Linked Data aims on making interlinked data available on the Web, and to interconnect data with the aim of increasing its value for users



A post on the Facebook group Reptile Road Mortality



Aims

- to transform the crowdsourced observation data to RDF model via using an ontology of citizens as sensors;
- to interlink the crowdsourced observation data with other Linked Data resources such as biodiversity (TaiCOL) and geospatial information (Geonames);
- to make the crowdsourced observation data
 - accessible to machines by using the Linked Data paradigm and
 - readable for humans by means of a faceted browser.

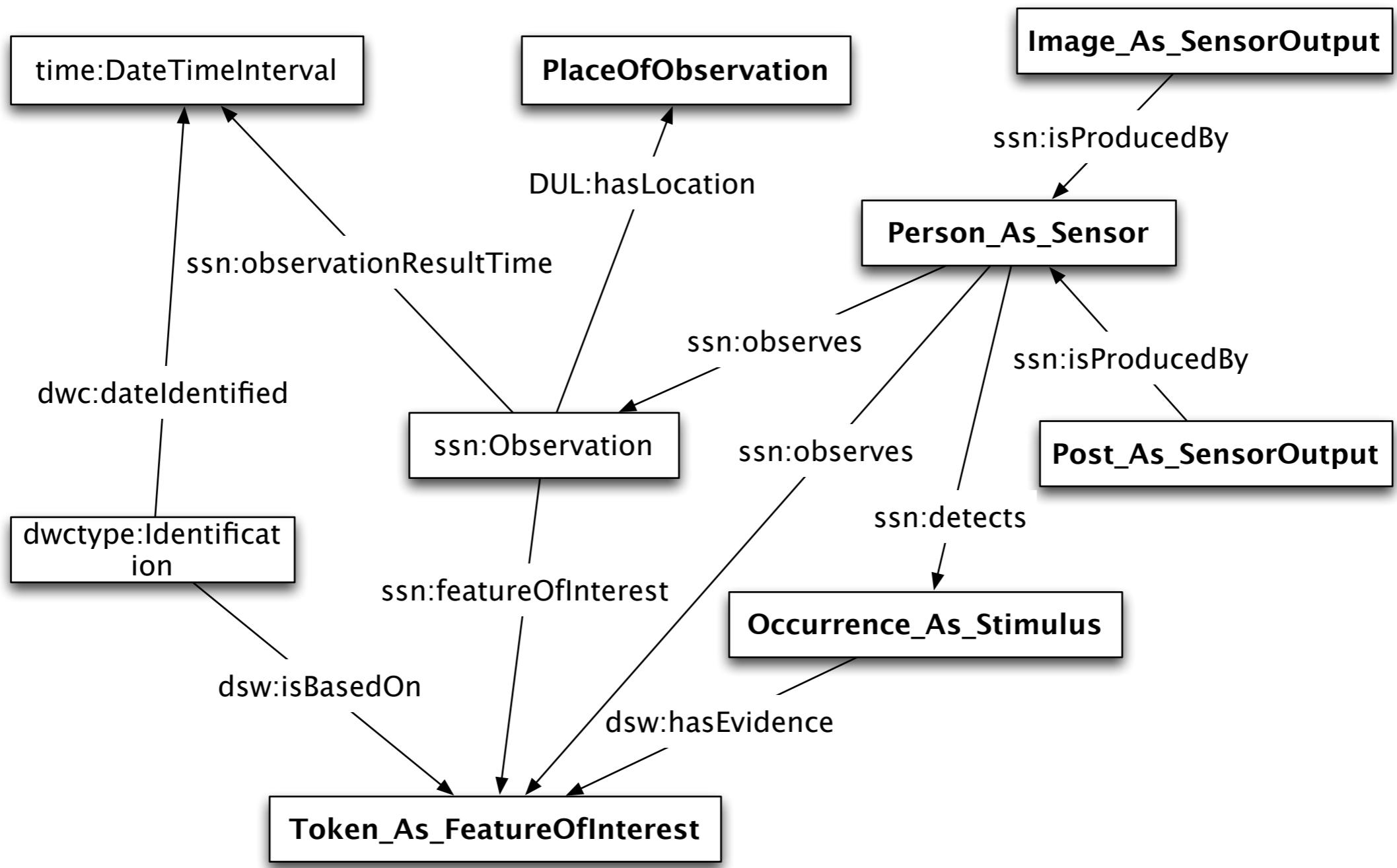
Information extraction

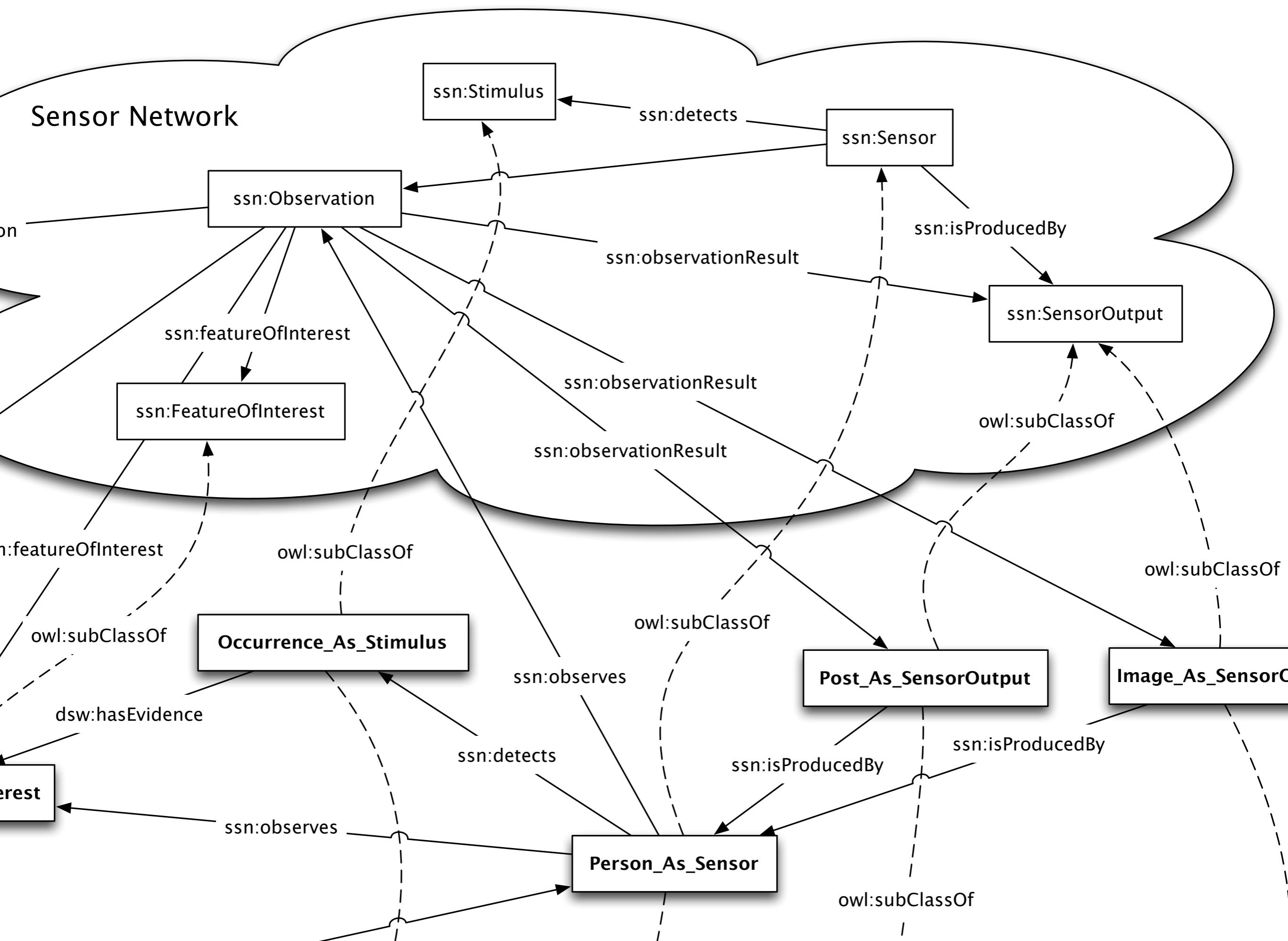
Named-Entity Recognition

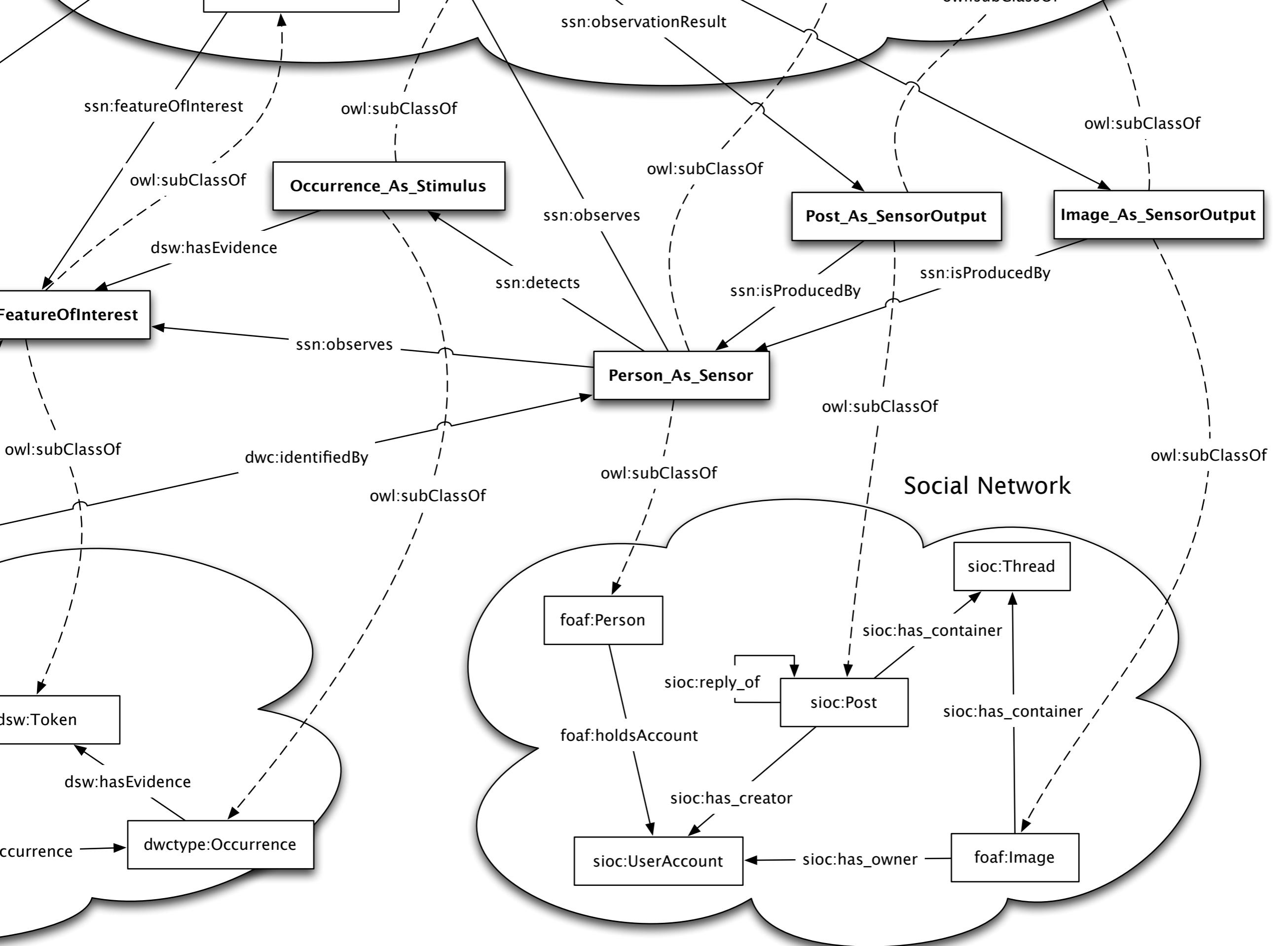
- Problems:
 - Chinese word segmentation
 - Chinese texts are character-based, not word-based
 - There is often no space between characters in written Chinese sentences
 - The lacks of domain-specific corpora such as geography and biodiversity.
 - Shorten names
- We compiled a geo-name lexicon from the Taiwan Geographic Names database and a species-name lexicon from the Taiwan Catalogue of Life databases (TaiCOL)
- The named-entity recognition approach we use was elaborated in a paper we previously published [10]

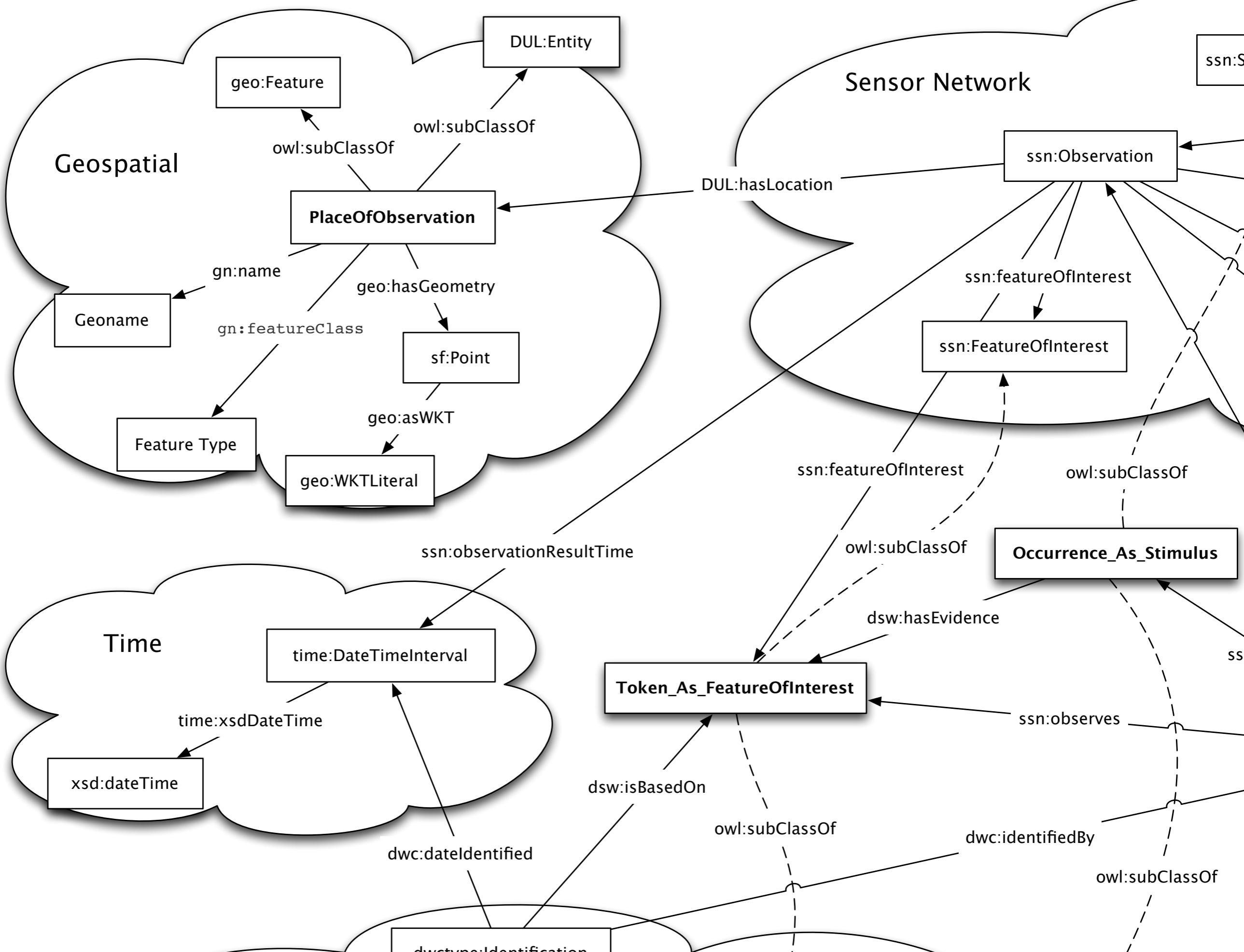
Ontology Engineering

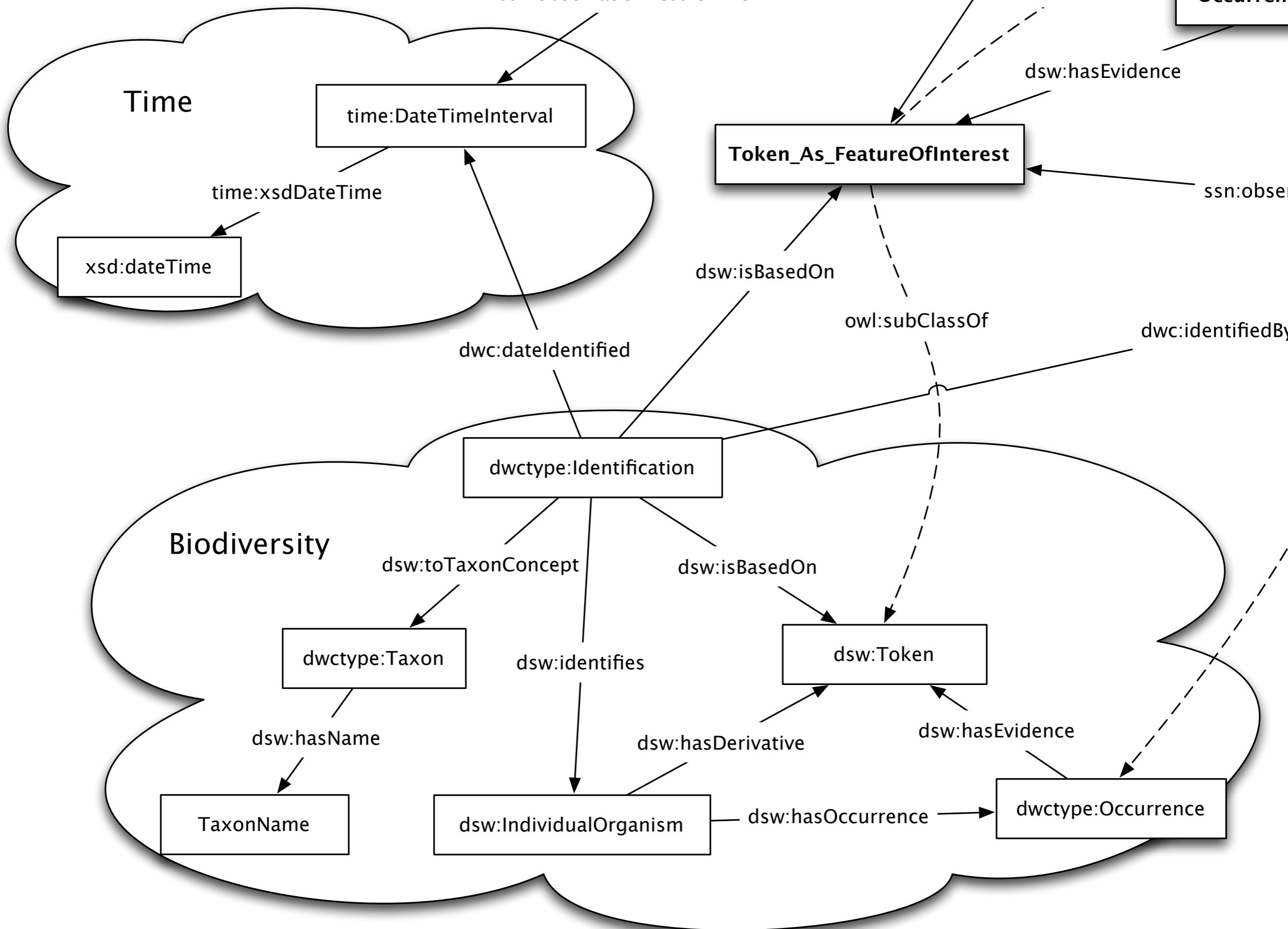
The ontology





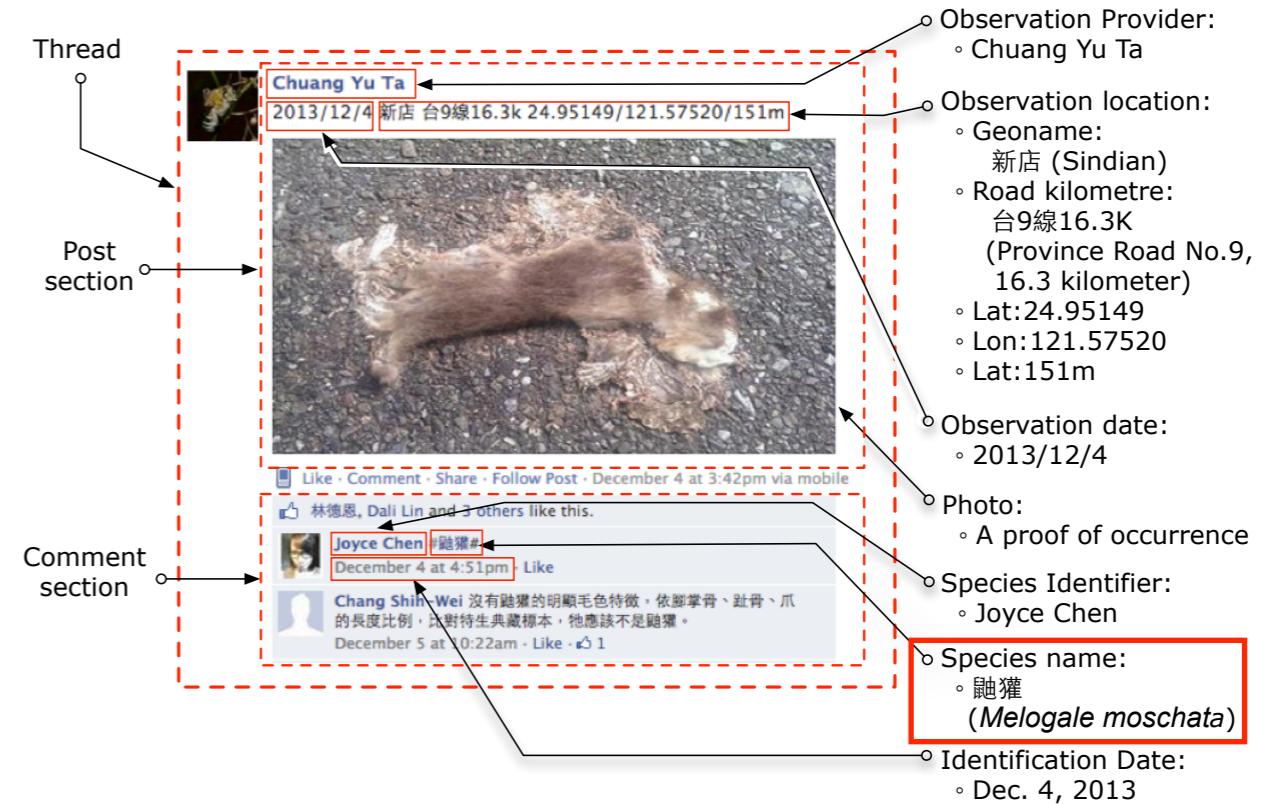






Transformations from crowdsourced data to RDF

Species names



```
eo:iden_559070840853748_01 rdf:type dwctype:Identification ,  
owl:NamedIndividual ;  
dwc:datIdentified eo:iden_time_559070840853748 ;  
dsw:identifies eo:idv_238918712815615_694835510557264 ;  
dsw:isBasedOn eo:token_559070840853748 ;  
dsw:toTaxonConcept taxon:380522 .
```

```
taxon:380522 rdf:type dwctype:Taxon ,  
owl:NamedIndividual ;  
dsw:hasName taibif:380522 ;  
skos:preLabel "Melogale moschata subaurantiaca" ;  
skos:altLabel "馯獾" .
```

The taxon concept of extract species name is linked to a URI in TaiBIF

Melogale moschata subaurantiaca (Swinhoe, 1862) at Taiwan Catalogue of Life LOD	
http://taibif.tw/lod/resource/Species/380522	
<hr/>	
Property	Value
skos:altLabel	<ul style="list-style-type: none">▪ Melogale moschata subaurantiaca
spv:altVernacularName_s	<ul style="list-style-type: none">▪ 臭獾 (zh-tw)▪ 鼬獾 (zh-tw)
spv:class	<ul style="list-style-type: none">▪ <http://taibif.tw/lod/resource/Class/4c42df70ba1a1173b20bfcc1132e7ed3>
spv:class_s	<ul style="list-style-type: none">▪ Mammalia▪ 哺乳綱 (zh-tw)
spv:family	<ul style="list-style-type: none">▪ <http://taibif.tw/lod/resource/Family/f22835e7f4d83d10625f718f21013407>
spv:family_s	<ul style="list-style-type: none">▪ Mustelidae▪ 貂科 (zh-tw)
spv:genus	<ul style="list-style-type: none">▪ <http://taibif.tw/lod/resource/Genus/5d4825579b45a29e247dd39d5deb29e8>
spv:genus_s	<ul style="list-style-type: none">▪ Melogale▪ 鼬獾屬 (zh-tw)
spv:kingdom	<ul style="list-style-type: none">▪ <http://taibif.tw/lod/resource/Kingdom/7e12691cc1761f6ffa845d697901d583>
spv:kingdom_s	<ul style="list-style-type: none">▪ Animalia▪ 動物界 (zh-tw)
rdfs:label	<ul style="list-style-type: none">▪ Melogale moschata subaurantiaca (Swinhoe, 1862)
spv:order	<ul style="list-style-type: none">▪ <http://taibif.tw/lod/resource/Order/ed06c3f53b30b788c742c29e6d6218ba>
spv:order_s	<ul style="list-style-type: none">▪ Carnivora▪ 食肉目 (zh-tw)
spv:parent	<ul style="list-style-type: none">▪ <http://taibif.tw/lod/resource/Genus/5d4825579b45a29e247dd39d5deb29e8>
spv:parentRankPath	<ul style="list-style-type: none">▪ Kingdom:Animalia/phylum:Chordata/class:Mammalia/order:Carnivora/family:Mustelidae/genus:Melogale
spv:phylum	<ul style="list-style-type: none">▪ <http://taibif.tw/lod/resource/Phylum/83a3673b63eeb827369a1e15e4185366>
spv:phylum_s	<ul style="list-style-type: none">▪ Chordata▪ 脊索動物門 (zh-tw)

The taxon name of extract species name is linked to a URI in TaiBIF

Melogale moschata subaurantiaca (Swinhoe, 1862) at Taiwan

Catalogue of Life LOD

<http://taibif.tw/lod/resource/ScientificName/380522>



Property	Value
spv:canonicalForm	<ul style="list-style-type: none"><http://taibif.tw/lod/resource/CanonicalForm/Melogale_moschata_subaurantiaca>
rdfs:label	<ul style="list-style-type: none">Melogale moschata subaurantiaca (Swinhoe, 1862)
is spv:scientificName of	<ul style="list-style-type: none"><http://taibif.tw/lod/resource/Species/380522>
rdf:type	<ul style="list-style-type: none">spv:ScientificName

This page shows information obtained from the SPARQL endpoint at <http://140.109.28.72:8890/sparql>.

[As Turtle](#) | [As RDF/XML](#) | [LOD.TW RDF Vizualizer](#)

Geographic names



```
eoe:placeOfOb_559070840853748 rdf:type eoe:PlaceOfObservation ,  
owl:NamedIndividual ;  
geo:hasGeometry eoe:point_559070840853748 ;  
gn:name "新店" ;  
owl:sameAs http://lod.tw/placenames/159624 .
```

```
eoe:point_559070840853748 rdf:type geo:Point ,  
owl:NamedIndividual ;  
w3c_geo:long "121.575200" ;  
w3c_geo:lat "24.951490" ;  
geo:asWKT "Point(121.575200 24.951490)"^^sf:wktLiteral .
```

The extract place name points to a URI in Taiwan Geographic Name Database

geoname_tw #159624 

Resource URI:
http://localhost:2020/resource/geoname_tw/159624

[Home](#) | [All geoname_tw](#)

Property	Value
tgn:Address	
tgn:County	新北市
tgn:Kind	地名
tgn:Name	新店
tgn:Phone_no	
tgn:Townname	新店區
tgn:Xcoord	121.53397E0 (xsd:decimal)
tgn:Ycoord	24.928386E0 (xsd:decimal)
tgn:en_hanyu	Xindian
tgn:en_tongyong	Sindian
tgn:id	159624 (xsd:integer)
rdfs:label	geoname_tw #159624
owl:sameAs	<gn:1674828>
rdf:type	tgn:geoname_tw

The server is configured to display only a limited number of values (limit per property bridge: 50).

Triplestore (BBN Parliament)

<http://lod.tw/parliament/>

The screenshot shows the 'Explore Repository' interface for the BBN Parliament triplestore. The top navigation bar includes links for Home, Operations (selected), Query, Explore, SPARQL/Update, Insert Data, Export, Indexes, and Admin. A search bar at the top allows entering a URI to start exploration. Below the search bar, the 'Graphs' section lists the Default Graph as <http://parliament.semwebcentral.org/parliament#MasterGraph>. The 'Classes and Properties' section is divided into two columns: 'Classes' and 'Properties'. The 'Classes' column lists various RDF classes such as <http://www.w3.org/2006/time#DateTimeInterval>, <http://www.opengis.net/ont/geosparql#Point>, <http://www.opengis.net/ont/geosparql#Feature>, etc. The 'Properties' column lists various RDF properties such as <http://www.opengis.net/ont/geosparql#hasGeometry>, <http://www.geonames.org/ontology#name>, <http://rs.tdwg.org/dwc/terms/identifiedBy>, etc.

Classes	Properties
http://www.w3.org/2006/time#DateTimeInterval	http://www.opengis.net/ont/geosparql#hasGeometry
http://www.opengis.net/ont/geosparql#Point	http://www.geonames.org/ontology#name
http://www.opengis.net/ont/geosparql#Feature	http://rs.tdwg.org/dwc/terms/identifiedBy
http://rs.tdwg.org/dwc/dwctype/Identification	http://rs.tdwg.org/dwc/terms/dateIdentified
http://purl.org/dsw/IndividualOrganism	http://purl.org/dsw/toTaxonConcept
http://purl.oclc.org/NET/ssnx/ssn#Observation	http://purl.org/dsw/IsBasedOn
http://lod.tw/ontologies/eoe.owl#Token_As_FeatureOfInterest	http://purl.org/dsw/hasName
http://lod.tw/ontologies/eoe.owl#TaxonName	http://purl.org/dsw/sf:asWKT
http://lod.tw/ontologies/eoe.owl#Post_As_SensorOutput	
http://lod.tw/ontologies/eoe.owl#PlaceOfObservation	
http://lod.tw/ontologies/eoe.owl#Person_As_Sensor	
http://lod.tw/ontologies/eoe.owl#Occurrence_As_Stimulus	
http://lod.tw/ontologies/eoe.owl#Occurrence_As_Stimulus	
http://lod.tw/ontologies/eoe.owl#Image_As_SensorOutput	
http://lod.tw/ontologies/eoe.owl#Image_As_SensorOutput	
http://lod.tw/ontologies/eoe.owl#Geoname	

Spatiotemporal SPARQL query

```
SELECT Distinct ?Obs ?POO_geo ?POO_wkt
WHERE{
    ?Obs a ssn:Observation;
    DUL:hasLocation ?POO ;
    ssn:observationResultTime ?Int .
    ?POO geo:hasGeometry ?POO_geo .
    ?POO_geo geo:asWKT ?POO_wkt .
    ?Int time:xsdDateTime ?Time_xsd .
    FILTER (geof:sfWithin(?POO_wkt,"POLYGON(
        121.756555 24.488236, 121.207238 24.488236,
        121.207238 25.141394, 121.756555 25.141394,
        121.756555 24.488236)"^^sf:wktLiteral))
    Filter (?Time_xsd > "2013-12-19T16:00:00Z"^^xsd:dateTime )
}
```

The query result of SPARQL query

Count: 3

Obs	POO_geo	POO_wkt
http://lod.tw/ontologies/eoe.owl#obs_790740334275056	http://lod.tw/ontologies/eoe.owl#point_790740334275056	"< http://www.opengis.net/def/crs/OGC/1.3/CRS84 >Point(121.341180 25.060235)" ^^< http://www.opengis.net/ont/sf#wktLiteral >
http://lod.tw/ontologies/eoe.owl#obs_10201138534854786	http://lod.tw/ontologies/eoe.owl#point_10201138534854786	"< http://www.opengis.net/def/crs/OGC/1.3/CRS84 >Point(121.509566 25.042220)" ^^< http://www.opengis.net/ont/sf#wktLiteral >
http://lod.tw/ontologies/eoe.owl#obs_736869536325952	http://lod.tw/ontologies/eoe.owl#point_736869536325952	"< http://www.opengis.net/def/crs/OGC/1.3/CRS84 >Point(121.365256 24.905200)" ^^< http://www.opengis.net/ont/sf#wktLiteral >

The faceted browser

<http://taibif.tw/vgd/ldow2014/viewer.php>

Class Tree

- └ http://www.w3.org/2002/07/owl#Thing
 - └ Animalia
 - └ Chordata
 - └ Aves
 - └ Reptilia
 - └ Squamata
 - └ Colubridae
 - └ Boiga
 - └ Boiga kraepelini Stejneger, 1895

Arbor Layout

(1)

Modify Your Query (5)

(2)

(3)

Observation Timeline

start

obs_790740334275056
obs_10201138534854786
obs_736869536325962

Sep 2013 Oct 2013 Nov 2013 Dec 2013 Jan 2014 Feb 2014 Mar 2014 Apr 2014 May

timeglider prev Dec 23, 2013 next

(4)

(5)

Conclusions

- This study explored the issues involved in the use of social media in citizen science projects, as well as reported our experiences in transferring unstructured collaborative information to structured data for scientific purposes.
- We shared our experiences in tackling the data collection from social process to scientific process.
- The successful implementation of this approach can further facilitate the development of social-media based citizen science projects.
- We believe it also has broader applications in user-generated content management, and promises to be a practical solution to an important design problem in citizen science projects on the Web.

Future work

- This study uses several tools for storing and visualizing the RDF triples. To make the browser more usable, a task to integrate the tools into a knowledge-based browser remains to be done in the future.
- Moreover, the triplified dataset should be considered for linkage to larger linked datasets such as DBpedia and other resources.



Thank you for your attentions!

Question?

dongpo@iis.sinica.edu.tw