



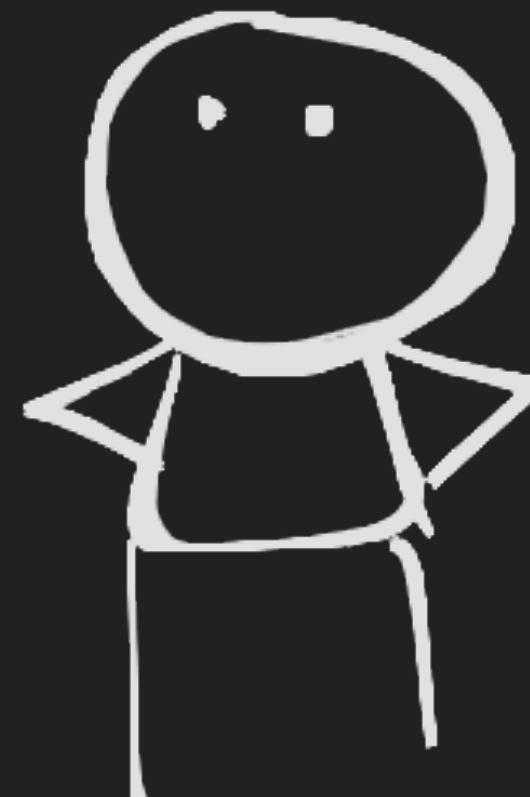
# Semantic search

Jérôme Gasperi

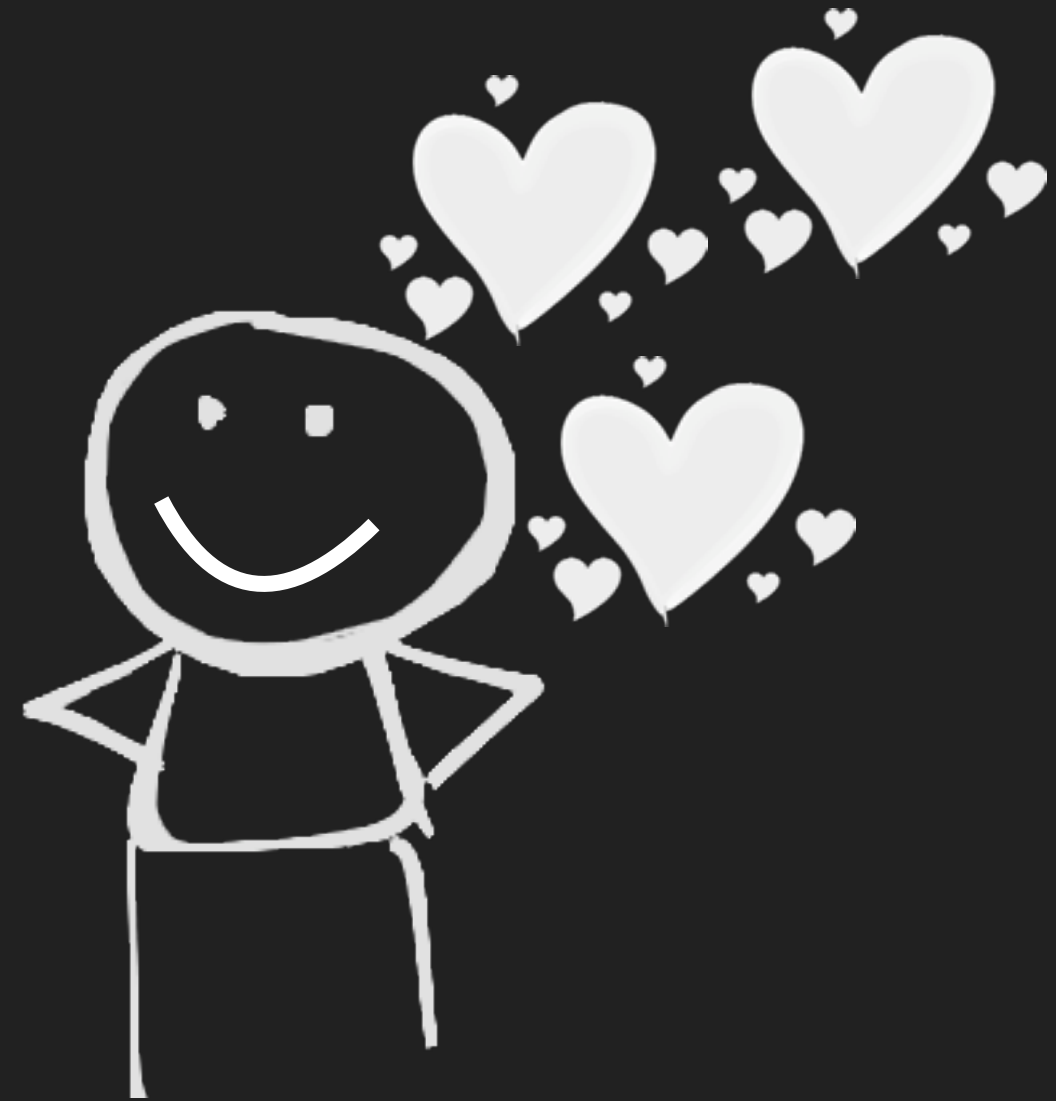
WGISS #37

Cocoa Beach, Florida - USA - April 16<sup>th</sup>, 2014

Semantic search helps users to  
find the **right data**



Semantic search helps users to  
find the **right data**



How to add semantics capabilities to EO products search services ?





Characterize products with **relevant** information.  
Think « users », not « experts »

1

Characterize products with **relevant** information.  
Think « users », not « experts »

2

Decode users natural language queries







Enrich products

Use **footprint** to enrich metadata from **exogenous** data



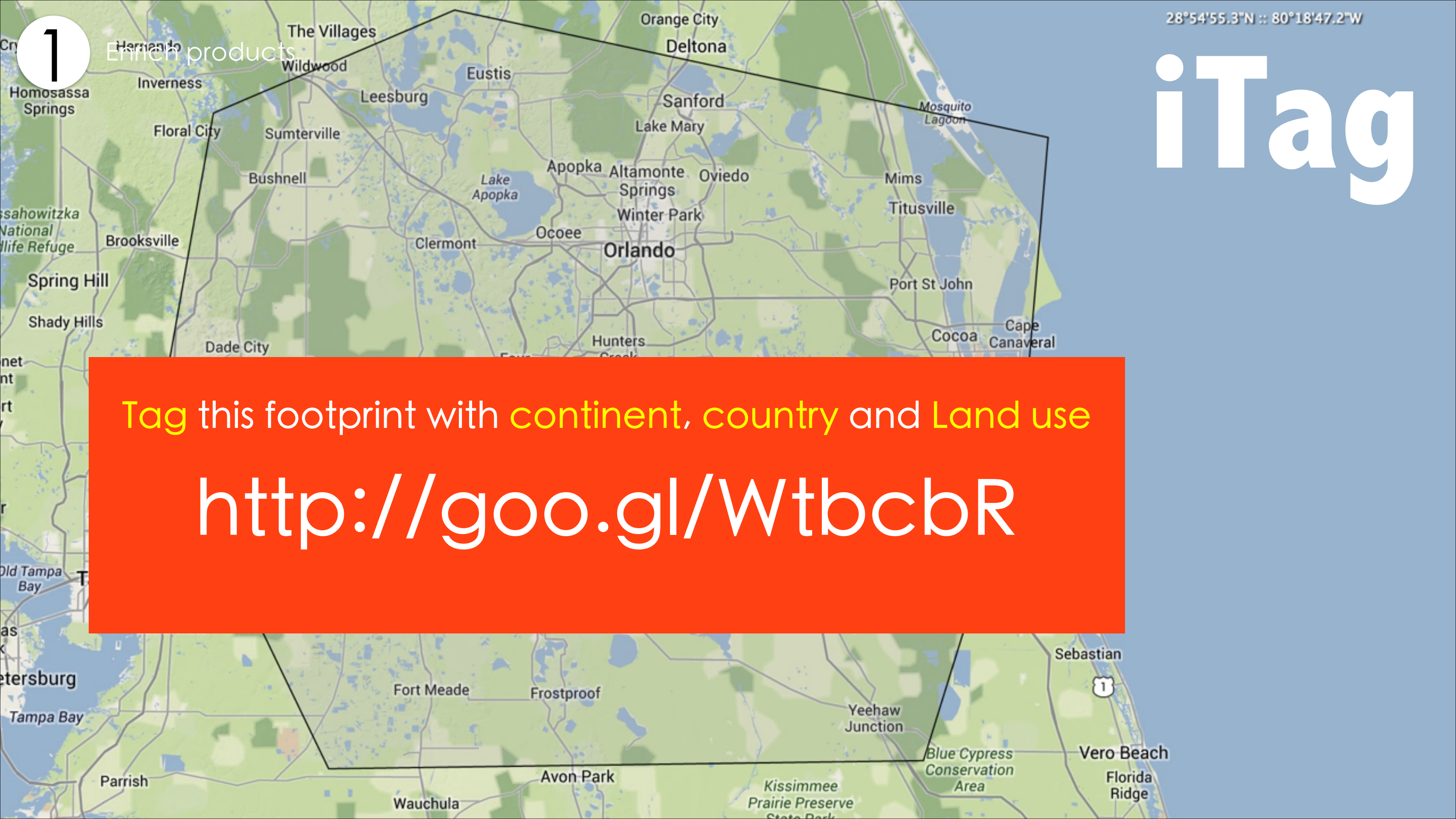
[github.com/jjrom/itag](https://github.com/jjrom/itag)

28°54'55.3"N :: 80°18'47.2"W

# iTag

1

Enrich products



Tag this footprint with **continent**, **country** and **Land use**

<http://goo.gl/WtbcbR>



2

Decode queries

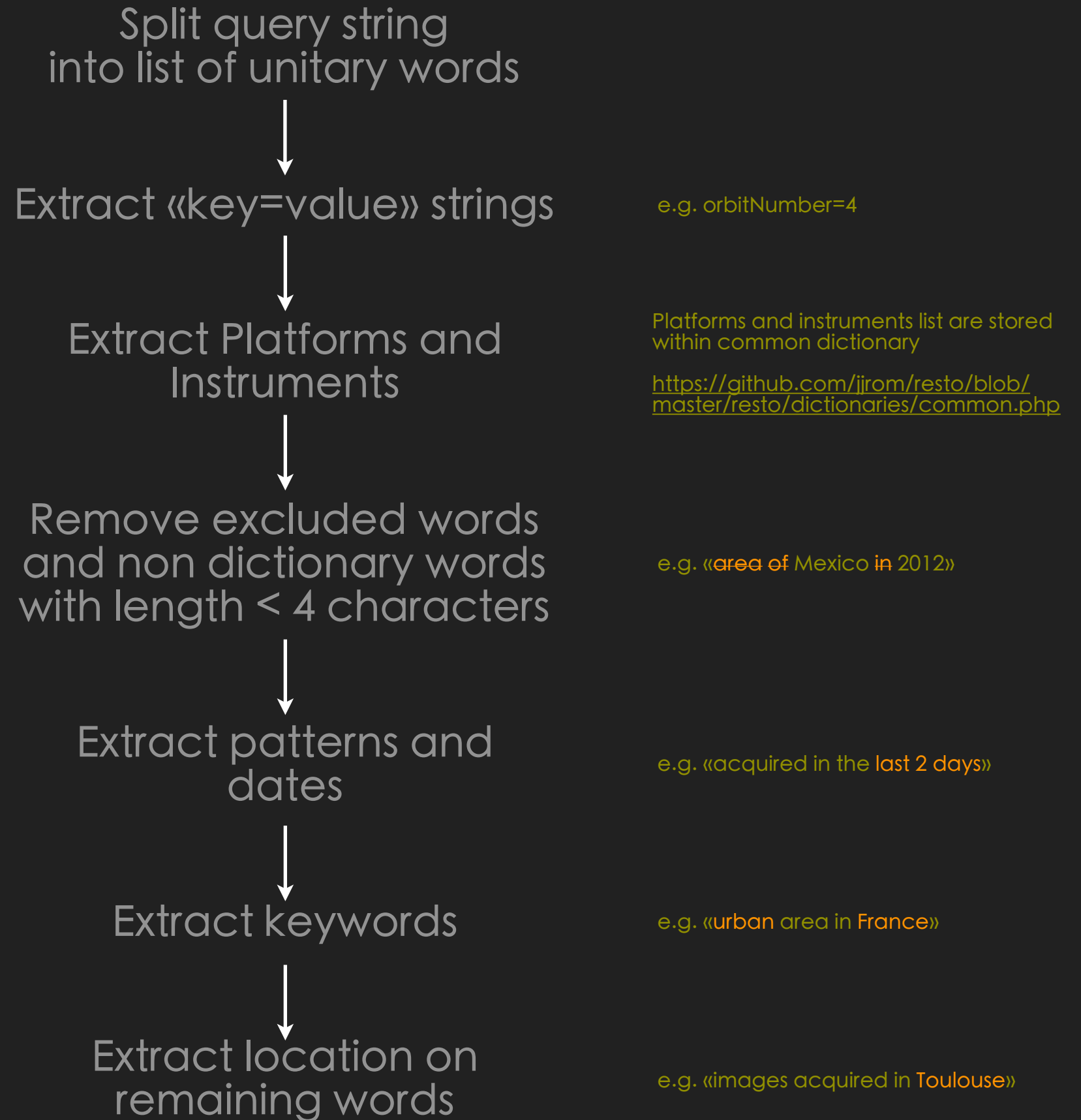
RESTo provides **semantic search** capabilities

It uses a **Query Analyzer** to translate query into a set of EO OpenSearch parameters

# 2

## Decode queries

Query string analysis algorithm is based on simple **recognition** of words and **patterns**



# 2

## Decode queries

### Recognized patterns

```
<with> "keyword"  
<without> "keyword"  
  
"quantity" <lesser> (than) "numeric" "unit"  
"quantity" <greater> (than) "numeric" "unit"  
"quantity" <equal> (to) "numeric" "unit"  
<lesser> (than) "numeric" "unit" (of) "quantity"  
<greater> (than) "numeric" "unit" (of) "quantity"  
<equal> (to) "numeric" "unit" (of) "quantity"  
"quantity" <between> "numeric" <and> "numeric" ("unit")  
<between> "numeric" <and> "numeric" "unit" (of) "quantity"  
  
<today>  
<yesterday>  
<before> "date"  
<after> "date"  
<between> "date" <and> "date"  
"numeric" "(year|day|month)" <ago>  
<last> "(year|day|month)"  
<last> "numeric" "(year|day|month)"  
"numeric" <last> "(year|day|month)"  
"(year|day|month)" <last>  
<since> "numeric" "(year|day|month)"  
<since> "month" "year"  
<since> "date"  
<since> "numeric" <last> "(year|day|month)"  
<since> <last> "numeric" "(year|day|month)"  
<since> <last> "(year|day|month)"  
<since> "(year|day|month)" <last>
```

# 2

## Decode queries

Words are stored within a dictionary

```
$dictionary = array(  
    'excluded' => array(  
        'than',  
        'image',  
        ...  
    ),  
    'modifiers' => array(  
        'ago' => 'ago',  
        'before' => 'before',  
        'after' => 'after',  
        ...  
    ),  
    'units' => array(  
        'm' => 'm',  
        'meter' => 'm',  
        'days' => 'days',  
        ...  
    ),  
    'numbers' => array(  
        'one' => '1',  
        ...  
    ),  
    'months' => array(  
        'january' => '01',  
        ...  
    ),  
    'quantities' => array(  
        'resolution' => 'resolution',  
        ...  
    ),  
    'keywords' => array(  
        'continent' => array(  
            'europe' => 'europe',  
            ...  
        )  
    )  
)
```

2

Decode queries

# Example

« Images of urban area in the US acquired in the last 10 days with less than 5 % of cloud cover »



2

Decode queries

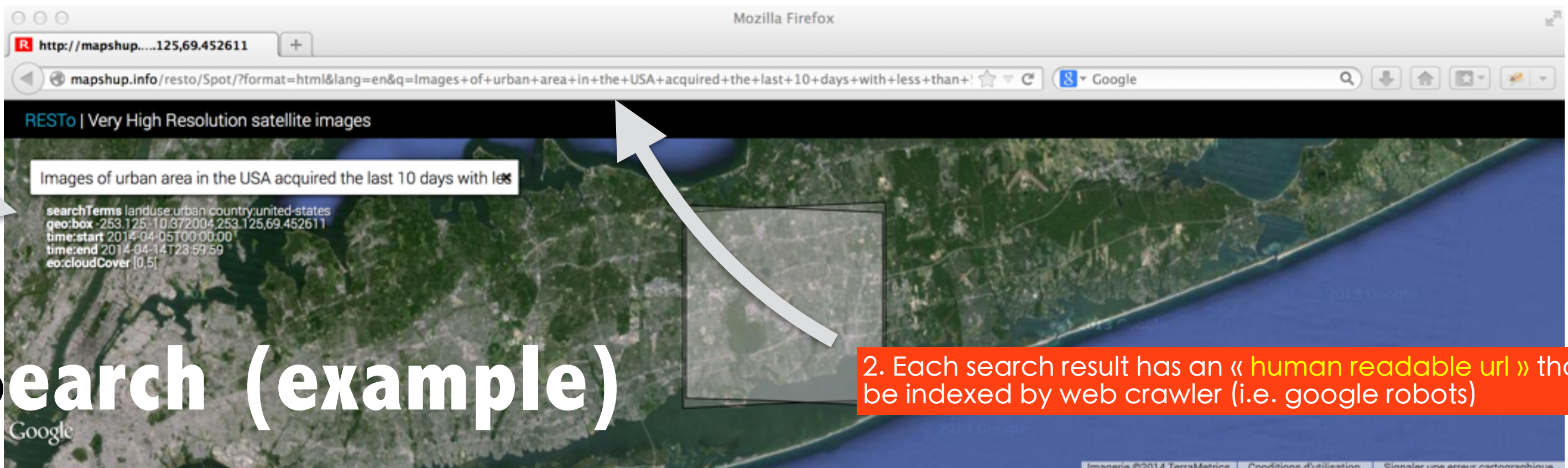
# Example

« Images of urban  
keyword

US  
location

last 10 days  
date

less 5% cloud cover »  
acquisition parameter



# Search (example)

1. Search parameters are derived from Natural Language query

2. Each search result has an « human readable url » that can be indexed by web crawler (i.e. google robots)

PHR1B   2014-04-13 15:59:03	PHR1B   2014-04-13 15:58:26	PHR1B   2014-04-13 15:58:01
Land cover Urban area (45%), Forest (37%)	Land cover Forest (30%), Urban area (25%), Water (23%)	Land cover Urban area (45%), Forest (38%)
Location United States, North America	Location United States, North America	Location United States, North America

3. Keywords on resources are links to search requests : they can be indexed by web crawler...and so on

RESTo | Very High Resolution satellite images

Images of urban area in the USA acquired the last 10 days with less

searchTerms landuse:urban country:united-states  
geo:box -253.125;10.372004,253.125,69.452611  
time:start 2014-04-05T00:00:00  
time:end 2014-04-14T23:59:59  
eo:cloudCover [0,5]

# Search (example)

# <http://goo.gl/GvMEHj>

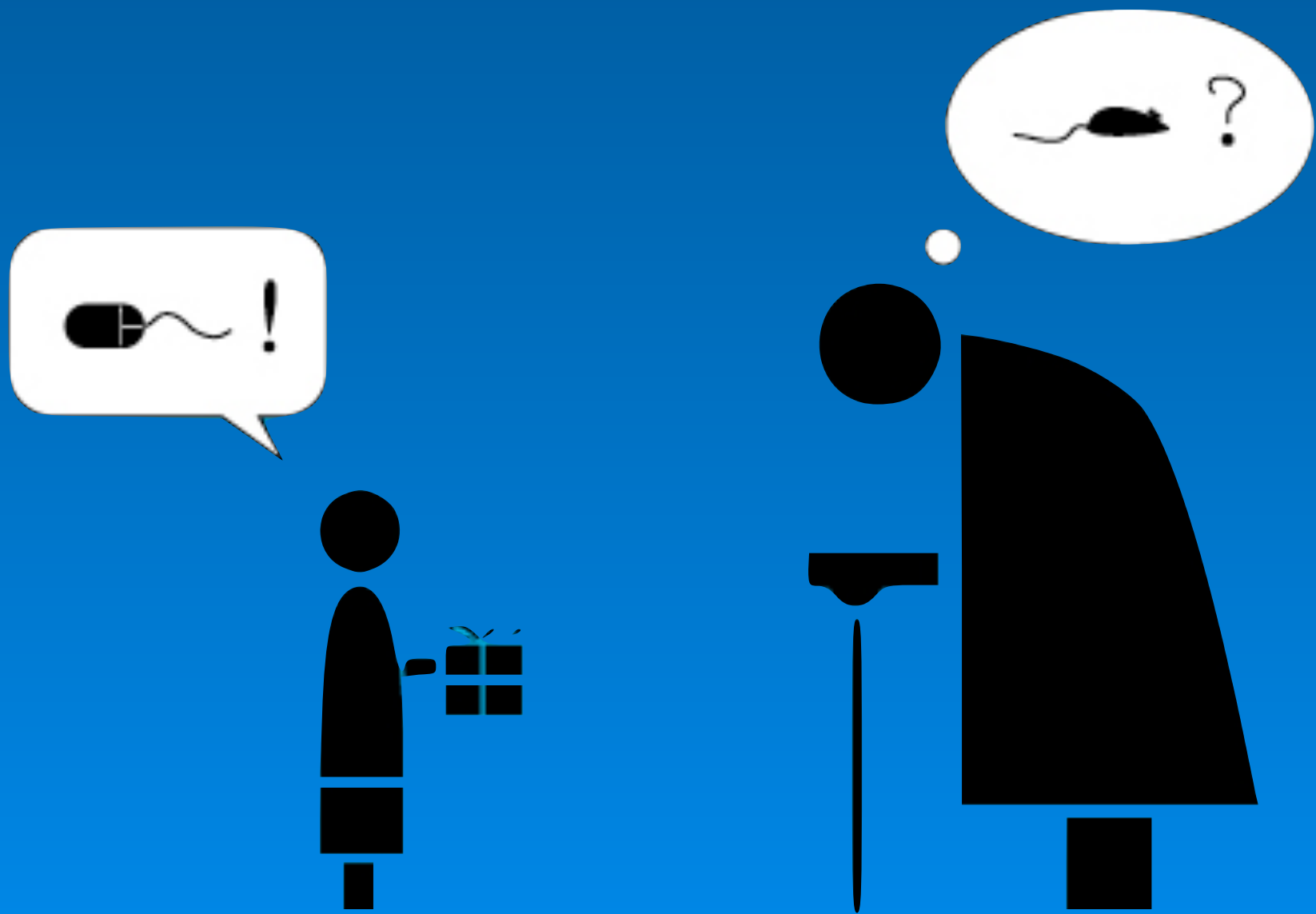
1. Search parameters are derived from Natural Language query

2. Each search result has an « human readable url » that can be indexed by web crawler (i.e. google robots)

PHR1B   2014-04-13 15:59:03	PHR1B   2014-04-13 15:58:26	PHR1B   2014-04-13 15:58:26	PHR1B   2014-04-13 15:58:01
Land cover Urban area (45%), Forest (37%)	Land cover Forest (30%), Urban area (25%), Water (23%)	Land cover Forest (28%), Urban area (25%), Water (25%)	Land cover Urban area (45%), Forest (38%)
Location United States, North America	Location United States, North America	Location United States, North America	Location United States, North America

3. Keywords on resources are links to search requests : they can be indexed by web crawler...and so on

Issues with keywords approach



Earthquakes in november 2008 in china

## Earthquakes in november 2008 in china

Ambiguous since it appears to be  
a location in New Zealand

« Linked data is the right way to do Semantic Web »

*Tim Berners-Lee*



Update RESTo JSON model to follow **JSON-LD** format

```
{
  "@context": "http://json-ld.org/contexts/person.jsonld",
  "@id": "http://dbpedia.org/resource/John_Lennon",
  "name": "John Lennon",
  "born": "1940-10-09",
  "spouse": "http://dbpedia.org/resource/Cynthia_Lennon"
}
```