The Kasabi Information Marketplace

knud.moeller@talis.com
@knudmoeller
19/04/2012, WWW2012, Lyon, France
A Place to...

- **publish** data
- **integrate** your data
- **monetize** your data
- **find** data
- **consume and use** data
• **web-based** platform
• **horizontal** marketplace
• RESTful APIs
• language bindings (Ruby, PHP, JS, Python)
• pytassium
What’s so special?

- Kasabi is based on **linked data** principles
  - data in **graph** structure (RDF)
  - **URIs** identify data items
  - data **links** to other datasets (context)
  - linked data **views**
What’s so special?

• Your data gets **APIs**
  - SPARQL endpoint
  - keyword search
  - lookup
  - reconciliation
  - custom APIs
Dashboard

My Dashboard

Activity
5th April 2012 - 19th April 2012

API key usage
Total credits used
6
View my usage report

My Datasets
Total requests
15
View my usage report

11 datasets with access granted
My API Key

0 users
on 3 datasets

My most accessed datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>world-geography</td>
<td>3</td>
</tr>
<tr>
<td>Dbpedia</td>
<td>2</td>
</tr>
<tr>
<td>Food</td>
<td>1</td>
</tr>
</tbody>
</table>

My datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Requests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Languages</td>
<td>15</td>
</tr>
</tbody>
</table>

Thursday 19 April 2012
Creating a dataset

Create Dataset

Logo:

Remove

Name:

WWW2012

Description:


Web page addresses and e-mail addresses turn into links automatically.

Allowed HTML tags: &lt;a&gt; &lt;em&gt; &lt;strong&gt; &lt;cite&gt; &lt;code&gt; &lt;ul&gt; &lt;ol&gt; &lt;li&gt; &lt;dl&gt; &lt;dt&gt; &lt;p&gt; &lt;h1&gt; &lt;h2&gt; &lt;h3&gt; &lt;h4&gt; Use [gist:####] where #### is your gist number to embed the gist.

More information about formatting options

Example Resource:

http://data.semanticweb.org/conference/www/2012/paper/10

http://data.semanticweb.org/person/tom-heath
<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>FR-R92B-58</td>
<td>Black</td>
<td>1059.31</td>
<td>1431.5</td>
<td>58</td>
<td>1016.04</td>
<td>3</td>
<td>6</td>
<td>00:00:00</td>
</tr>
<tr>
<td>2</td>
<td>FR-R92R-58</td>
<td>Red</td>
<td>1059.31</td>
<td>1431.5</td>
<td>58</td>
<td>1016.04</td>
<td>6</td>
<td></td>
<td>00:00:00</td>
</tr>
<tr>
<td>3</td>
<td>HL-U509-R</td>
<td>Red</td>
<td>13.0863</td>
<td>34.99</td>
<td></td>
<td></td>
<td>33</td>
<td></td>
<td>00:00:00</td>
</tr>
<tr>
<td>4</td>
<td>HL-U509</td>
<td>Black</td>
<td>13.0863</td>
<td>34.99</td>
<td></td>
<td></td>
<td>33</td>
<td></td>
<td>00:00:00</td>
</tr>
<tr>
<td>5</td>
<td>SO-B909-M</td>
<td>White</td>
<td>3.3963</td>
<td>9.5 M</td>
<td></td>
<td></td>
<td>27</td>
<td>18</td>
<td>00:00:00</td>
</tr>
<tr>
<td>6</td>
<td>SO-B909-L</td>
<td>White</td>
<td>3.3963</td>
<td>9.5 L</td>
<td></td>
<td></td>
<td>27</td>
<td>18</td>
<td>00:00:00</td>
</tr>
<tr>
<td>7</td>
<td>HL-U509-B</td>
<td>Blue</td>
<td>13.0863</td>
<td>34.99</td>
<td></td>
<td></td>
<td>36</td>
<td>33</td>
<td>00:00:00</td>
</tr>
<tr>
<td>8</td>
<td>CA-1098</td>
<td>Multi</td>
<td>6.9223</td>
<td>8.99</td>
<td></td>
<td></td>
<td>23</td>
<td>2</td>
<td>00:00:00</td>
</tr>
<tr>
<td>9</td>
<td>LJ-0192-S</td>
<td>Multi</td>
<td>38.4923</td>
<td>49.99 S</td>
<td></td>
<td></td>
<td>25</td>
<td>11</td>
<td>00:00:00</td>
</tr>
<tr>
<td>10</td>
<td>LJ-0192-M</td>
<td>Multi</td>
<td>38.4923</td>
<td>49.99 M</td>
<td></td>
<td></td>
<td>25</td>
<td>11</td>
<td>00:00:00</td>
</tr>
<tr>
<td>11</td>
<td>LJ-0192-L</td>
<td>Multi</td>
<td>38.4923</td>
<td>49.99 L</td>
<td></td>
<td></td>
<td>25</td>
<td>11</td>
<td>00:00:00</td>
</tr>
<tr>
<td>12</td>
<td>LJ-0192-X</td>
<td>Multi</td>
<td>38.4923</td>
<td>49.99 XL</td>
<td></td>
<td></td>
<td>25</td>
<td>11</td>
<td>00:00:00</td>
</tr>
<tr>
<td>13</td>
<td>FR-R92R-62</td>
<td>Red</td>
<td>868.6342</td>
<td>1431.5</td>
<td>62</td>
<td>1043.26</td>
<td>13</td>
<td>6</td>
<td>00:00:00</td>
</tr>
<tr>
<td>14</td>
<td>FR-R92R-44</td>
<td>Red</td>
<td>868.6342</td>
<td>1431.5</td>
<td>44</td>
<td>961.61</td>
<td>13</td>
<td>6</td>
<td>00:00:00</td>
</tr>
<tr>
<td>15</td>
<td>FR-R92R-48</td>
<td>Red</td>
<td>868.6342</td>
<td>1431.5</td>
<td>48</td>
<td>979.75</td>
<td>13</td>
<td>6</td>
<td>00:00:00</td>
</tr>
<tr>
<td>16</td>
<td>FR-R92R-52</td>
<td>Red</td>
<td>868.6342</td>
<td>1431.5</td>
<td>52</td>
<td>997.9</td>
<td>13</td>
<td>6</td>
<td>00:00:00</td>
</tr>
<tr>
<td>17</td>
<td>FR-R92R-56</td>
<td>Red</td>
<td>868.6342</td>
<td>1431.5</td>
<td>56</td>
<td>1016.04</td>
<td>13</td>
<td>6</td>
<td>00:00:00</td>
</tr>
<tr>
<td>18</td>
<td>FR-R38B-58</td>
<td>Black</td>
<td>204.6251</td>
<td>337.22</td>
<td>58</td>
<td>1115.83</td>
<td>18</td>
<td>9</td>
<td>00:00:00</td>
</tr>
<tr>
<td>19</td>
<td>FR-R38B-60</td>
<td>Black</td>
<td>204.6251</td>
<td>337.22</td>
<td>60</td>
<td>1124.9</td>
<td>18</td>
<td>9</td>
<td>00:00:00</td>
</tr>
<tr>
<td>20</td>
<td>FR-R38B-62</td>
<td>Black</td>
<td>204.6251</td>
<td>337.22</td>
<td>62</td>
<td>1133.98</td>
<td>18</td>
<td>9</td>
<td>00:00:00</td>
</tr>
<tr>
<td>21</td>
<td>FR-R38R-44</td>
<td>Red</td>
<td>187.1571</td>
<td>337.22</td>
<td>44</td>
<td>1052.33</td>
<td>18</td>
<td>9</td>
<td>00:00:00</td>
</tr>
<tr>
<td>22</td>
<td>FR-R38R-48</td>
<td>Red</td>
<td>187.1571</td>
<td>337.22</td>
<td>48</td>
<td>1070.47</td>
<td>18</td>
<td>9</td>
<td>00:00:00</td>
</tr>
<tr>
<td>23</td>
<td>FR-R38R-52</td>
<td>Red</td>
<td>187.1571</td>
<td>337.22</td>
<td>52</td>
<td>1088.62</td>
<td>18</td>
<td>9</td>
<td>00:00:00</td>
</tr>
<tr>
<td>24</td>
<td>FR-R38R-58</td>
<td>Red</td>
<td>187.1571</td>
<td>337.22</td>
<td>58</td>
<td>1115.83</td>
<td>18</td>
<td>9</td>
<td>00:00:00</td>
</tr>
</tbody>
</table>
CSV2RDF Conversion

<#weight> a :Resource ;
   :identity [ :source_column 8 ;
   :process ( :regex ) ;
   :regex_match "^(.+)$" ;
   :regex_output "${1}gr" ;
]
   :type gr:QuantitativeValueFloat ;
   :attribute [ :property gr:hasValue ; :source_column 8 ; :datatype xsd:float ] ,
   [ :property gr:hasUnitOfMeasurement ; :value "GRM" ; :datatype xsd:string ] .

<#category> a :Resource ;
   :identity [ :source_column 10 ;
]
   :type owl:Class ;
CSV2RDF Conversion

```
gr:hasValue "1006.97"^^xsd:float ;
a gr:QuantitativeValueFloat .

<http://data.kasabi.com/dataset/adventure_works/weights/1006.97gr>
gr:hasUnitOfMeasurement "GRM"^^xsd:string ;
gr:hasValue "1006.97"^^xsd:float ;
a gr:QuantitativeValueFloat .

<http://data.kasabi.com/dataset/adventure_works/weights/1016.04gr>
gr:hasUnitOfMeasurement "GRM"^^xsd:string ;
gr:hasValue "1016.04"^^xsd:float ;
a gr:QuantitativeValueFloat .

<http://data.kasabi.com/dataset/adventure_works/weights/1025.11gr>
gr:hasUnitOfMeasurement "GRM"^^xsd:string ;
gr:hasValue "1025.11"^^xsd:float ;
a gr:QuantitativeValueFloat .
```

https://github.com/mmmmmmrrob/Vertere-RDF

Thursday 19 April 2012
Datasets

Metadata for the **21st World Wide Web Conference**, April 2012, Lyon, France, including authors, papers, organisations, etc.

More extensive conference data can be found at [SWDF](http://kasabi.com/dataset/www2012).

Published: 19th April 2012
Updated: 19th April 2012
Resources: 2920

Explore the dataset

View the [vocabularies and classes](http://kasabi.com/dataset/www2012) used in the dataset, or see some [sample resources](http://kasabi.com/dataset/www2012).

Read the [Developer Documentation](http://kasabi.com/dataset/www2012) or [edit it](http://kasabi.com/dataset/www2012).

Browse as [Linked Data](http://kasabi.com/dataset/www2012).
**Dataset Description**


### WWW2012

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Dataset</td>
</tr>
<tr>
<td>Title</td>
<td>WWW2012</td>
</tr>
<tr>
<td>Description</td>
<td>The WWW2012 dataset</td>
</tr>
<tr>
<td>Homepage</td>
<td>www2012</td>
</tr>
<tr>
<td>Date created</td>
<td>2012-04-18T23:39:47+00:00</td>
</tr>
<tr>
<td>Date modified</td>
<td>2012-04-19T00:35:48+00:00</td>
</tr>
<tr>
<td>Distinct subjects</td>
<td>2920</td>
</tr>
<tr>
<td>Status endpoint</td>
<td>status</td>
</tr>
<tr>
<td>Jobs endpoint</td>
<td>jobs</td>
</tr>
<tr>
<td>Attribution endpoint</td>
<td>attribution</td>
</tr>
<tr>
<td>Store endpoint</td>
<td>store</td>
</tr>
<tr>
<td>Example resource</td>
<td>1042 tom-heath</td>
</tr>
<tr>
<td></td>
<td><a href="http://xmlns.com/foaf/0.1/">http://xmlns.com/foaf/0.1/</a></td>
</tr>
</tbody>
</table>
Dataset Description

Dataset Description


```json
{
  "value": "The WWW2012 dataset",
],
  "http://xmlns.com/foaf/0.1/homepage": [
    {
      "type": "uri",
      "value": "http://kasabi.com/dataset/www2012"
    }
  ],
  "http://purl.org/dc/terms/created": [
    {
      "type": "literal",
      "value": "2012-04-18T23:39:47+00:00"
    }
  ],
  "http://purl.org/dc/terms/modified": [
    {
      "type": "literal",
      "value": "2012-04-19T00:35:48+00:00"
    }
  ],
  "http://rdfs.org/ns/void#distinctSubjects": [
  
}
```
APIs

Default APIs

**SPARQL Endpoint**

Use the SPARQL 1.1 query language to perform structured queries against a dataset. Useful for performing precise queries against a dataset whose structure you understand.

**Sample queries**

- Papers about "Online Communities"  
- All Paper Subjects

**Contributed APIs**

- Paper by Topic
  - SPARQL Stored Procedure

Create a sample query
Create an API
SPARQL API


API Key: 
Output: json

SPARQL Query

PREFIX swrc: <http://swrc.ontoware.org/ontology#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>

SELECT DISTINCT ?subject
WHERE {
  ?paper a swrc:InProceedings ;
  dc:subject ?subject .
}
ORDER BY ?subject

Tips
Use the SPARQL API to match patterns in the graph to extract data

Click to add prefix to query
- Bibliographic Ontology
- Dublin Core Elements
- Dublin Core Terms
- FOAF
- W3C Geo Schema
- Music Ontology
- Programmes Ontology
- ORG
- OWL
- RDF Namespace
- RDF Schema
- Relationship
- Reviews
- RSS 1.0
- SIOC Core Ontology
- SKOS Core
- VCard
- XML Schema

Read the API Docs for more information.

Thursday 19 April 2012
Search API

**Keyword Search**


- API Key: 
- Query: linked
- Sort criteria: 
- Number of results: 10
- Offset: 0
- Output: json

**Tips**
Read the API Docs for more information.

**Request**

**Response**

**RESPONSE BODY**

```json
{
    "head": {
        "query": "linked",
        "startIndex": 0,
        "totalResults": 14
    },
    "results": [
        {
            "url": "http://data.semanticweb.org/conference/www/2012/phd/26",
            "title": "From Linked Data to Linked Entities: A Migration Path",
            "score": 3.0511758
        },
        {
            "url": "http://data.semanticweb.org/conference/www/2012/dev/17",
            "title": "LDF - A Framework for Large-Scale Linked Data Integration"
        }
    ]
}
```
Lookup API Explorer

Use this form to test out the Lookup API for the WWW2012 dataset


API Key: 

Resource URI: http://data.semanticweb.org/person/tom-heath

Output: json

Look up

Response

RESPONSE BODY

{

  "http://data.semanticweb.org/person/tom-heath": {
    "http://swrc.ontoware.org/ontology#affiliation": [
      {
        "value": "http://data.semanticweb.org/organization/talis-education-ltd",
        "type": "uri"
      },
      {
        "value": "http://data.semanticweb.org/organization/talis-information-limited",
        "type": "uri"
      }
    ],
    "http://xmlns.com/foaf/0.1/name": [
      {
        "value": "Tom Heath",
        "type": "literal"
      }
    ]
  }
}
Reconciliation API

Reconciliation API Explorer

Use this form to test out the Reconciliation API for the WWW2012 dataset


API Key: [input field]
Query: [input field] Tom Heath
Type URI: [input field]
Limit: [input field] 3

Reconcile

Tips
The reconciliation API allows labels and simple identifiers to be looked up in a dataset to find the URI of an item. This is useful to link together datasets.

The API is supported in Google Refine so can be used when tidying up data using that tool.

Read the API Docs for more information.

Response Body

```json
{
  "result": [
    {
      "id": "http://data.semanticweb.org/person/tom-heath",
      "name": "Tom Heath",
      "type": [
        "http://xmlns.com/foaf/0.1/Person"
      ],
      "score": 1,
      "match": true
    }
  ]
}
```
Create an API in WWW2012

You can create the following types of services within this dataset

Create SPARQL Stored Procedure

This type of API allows you to bind a SPARQL query to a URL so it will be automatically executed on request. Parameters can be passed from the query string and transformations can be applied to create custom data formats.

You’ll need:

- A SPARQL query that extracts the data. You can specify how parameters can be included from the request URL.
- Stored Procedures have a default limit of 10 results. This can be changed by using the limit keyword in your SPARQL query.
- Optionally, an XSLT transform to create custom output formats, but XML and JSON is available by default. We’ve created some generic ones you can use.

Setting up the API is quite easy and shouldn’t take more than a few minutes. Read the SPARQL Stored Procedure documentation for more details.

Create Linked Data API

A Linked Data API allows you to define a custom RESTful API for extracting data from a dataset by defining the graph patterns, e.g. types of entities and their relationships, that are of interest.

You’ll need:

- familiarity with Turtle for creating and editing the configuration file
- a good working knowledge of the Linked Data API vocabulary and/or a starter template that you can customize

Read the Linked Data API documentation to find pointers to get you started.
Custom APIs

Query:

```sql
PREFIX swrc: <http://swrc.ontoweb.org/swrc/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?paper ?title
WHERE {
    ?paper a swrc:InProceedings;
    rdfs:label ?title ;
    dc:subject ?subject .
}
```

Parameters:

- Parameter: 
  - subject

Default Value:

- AJAX

Required:

- Optional

Type:

- literal

Base:

There is more...

- [http://kasabi.com/doc/api](http://kasabi.com/doc/api)

- data management APIs (update, jobs, status, ...)

Thursday 19 April 2012
Does it cost anything?

Worker Bee

Best for individuals

£15/mo

All the features of the Free Bee plan with:
- 60,000 credits per month
- publish up to 20 datasets
- access to premium data
- bulk data loading
- own domain for data pages
- email support

Sign Up

Swarm

Best for teams

£120/mo

All the features of the Worker Bee plan with:
- 400,000 credits per month
- publish up to 100 datasets
- set royalty rate for data
- brandable data pages
- statistics API

Sign Up

Hive

Best for enterprise

£1500/mo

All the features of the Swarm plan with:
- 5,000,000 credits per month
- publish up to 500 datasets
- prioritised data loading
- custom licenses
- private datasets
- telephone support
- service level agreement

Sign Up

Free Bee

Explore the benefits of Kasabi

FREE

Sign Up

9,000 credits per month

access any standard datasets

publish up to 5 datasets

standard licenses

browsable data pages

data update API

live data updates

data usage statistics

Thursday 19 April 2012
Summary

- Kasabi is a platform to **publish, link, find** and **consume** data
- based on linked data principles
- Linked Data as a Service
- **APIs** over your data
- data in different **flavours** (turtle, json, rdf/xml)
Keep in touch!

- http://kasabi.com
- http://blog.kasabi.com/
- Twitter: @kasabi
- IRC: #kasabi (freenode.net)
- this presentation:
  http://www.slideshare.net/dunken69/the-kasabi-information-marketplace

This work is licensed under a Creative Commons Attribution 3.0 Unported License.
Under the Hood

• **Cohodo**

• *(used to be Talis Platform)*

• **distributed data platform**

• **load balancing, data replication, etc.**