

# Hosting Queryable and Highly Available Linked Data for Free

Luca Matteis *and* Ruben Verborgh

<http://luca.org> / <http://ruben.verborgh.org/>

[@lmatteis](#) / [@RubenVerborgh](#)

Bioversity International / Multimedia Lab --- Ghent University --- iMinds

# Motivation 1

SPARQL endpoints require the need to buy and configure complex servers.

You need to worry about:

- having the funds to keep the server running
- making sure the system is up-to-date
- many other sys-admin tasks

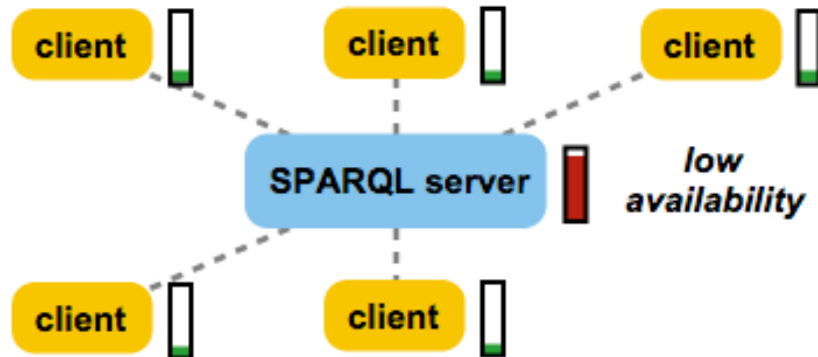
## *TL;DR*

Hosting an RDF file is a lot easier than hosting a SPARQL endpoint.

## Motivation 2

SPARQL endpoints *suffer from low-availability*.

They offer consumers the ability to run any query they want. Consumers will run any query they want and will quickly overload the server with too many complex requests.



## Motivation 3

Hosting code on version control systems (GitHub, BitBucket) or hosting data on cloud storages (Google Drive, Dropbox) is usually **free**.

Why isn't hosting Linked Data also free?

# Problem

Hosting SPARQL endpoints requires too much effort, both in terms of cost and server maintenance.

Fortunately SPARQL endpoints are not the only way of publishing queryable Linked Data.

*Triple Pattern Fragments* is a way of publishing queryable Linked Data on low cost servers.

# LDstatic & LDF-GAE

We have developed two tools, LDstatic & LDF-GAE, that implement the *TPF* protocol on low cost servers.

## ***TL;DR***

Users can run SPARQL queries against Linked Data published on online file repositories and cloud hosting services such as **GitHub**, **Google Code**, **Google App Engine** or **Dropbox**.

# LDstatic



**GitHub Pages**, and other online file repositories, can serve static HTML files.

It can also serve other type of content, such as N-Triples (`.nt`).

This means that GitHub Pages can serve triple pattern fragments.

# LDstatic

We want to match

`<foo> <bar> "literal" .`

using:

`/?subject=<foo>`

`/?predicate=<bar>`

`/?object="literal"`

`/?subject=<foo>&predicate=<bar>`

...



We need at least 8 static files for each triple if we want to match all combinations.

# LDstatic

Having all combinations is important for triple pattern fragments enabled clients so they can run complex queries, even SPARQL queries!

We can host queryable linked data on GitHub Pages:

<http://lmatteis.github.io/ldstatic/>

# LDF-GAE



We now talk about our other TPF implementation, which lets us run SPARQL queries on [Google App Engine](http://ldf-gae.appspot.com/).

<http://ldf-gae.appspot.com/>

# LDF-GAE

Storing triples on App Engine's high-replication datastore is quite simple. We store them using *Apache Jena's* representation.

Stored as:

```
{  
  "subject"   : triple.getSubject().toString(),  
  "predicate": triple.getPredicate().toString(),  
  "object"    : triple.getObject().toString()  
}
```

# LDF-GAE

To retrieve each triple pattern combination we use App Engine's native API calls:

```
// query datastore
var query = new Query('triple');
query.addFilter('subject',
    Query.FilterOperator.EQUAL, request.getParameter('subject'));
query.addFilter('predicate',
    Query.FilterOperator.EQUAL, request.getParameter('predicate'));
query.addFilter('object',
    Query.FilterOperator.EQUAL, request.getParameter('object'));
```

# LDF-GAE

Here's what a request looks like:

```
$curl http://ldf-gae.appspot.com/↵
```

```
?predicate=↵
```

```
http%3A%2F%2Fxmlns.com%2Ffoaf%2F0.1%2Fname
```

## Response:

```
<http://www.kjetil.kjernsmo.net/foaf#me> <http://xmlns.com/foaf/0.1/name> "Kjetil Kjernsmo" .  
<http://www.cs.umd.edu/~hendler/2003/foaf.rdf#jhendler> <http://xmlns.com/foaf/0.1/name> "Jim Hendler" .  
<http://www.w3.org/People/Jacobs/contact.rdf#IanJacobs> <http://xmlns.com/foaf/0.1/name> "Ian Jacobs" .  
<http://id.ecs.soton.ac.uk/person/60> <http://xmlns.com/foaf/0.1/name> "Les Carr" .  
<http://people.csail.mit.edu/psz/foaf.rdf#me> <http://xmlns.com/foaf/0.1/name> "Peter Szolovits" .  
<http://www.w3.org/People/EM/contact#me> <http://xmlns.com/foaf/0.1/name> "Eric Miller" .
```

## How is this novel?

We can exploit a variety of **free** hosting services to publish Linked Data and have clients query them. No more sys-admin tasks. Less down times.

For small datasets this is **more sustainable, more economical** and best of all it's **easier!**

# How to use?

Both tools are on GitHub with documentation and you can use them right now:

<https://github.com/lmatteis/ldstatic>

<https://github.com/lmatteis/ldf-gae>



# Conclusion

We proved that triple pattern fragments is a type of linked data can be published with high availability on low-cost Web servers ranging from basic file repositories to highly scalable cloud platforms.

# References

Linked Data Fragments --

<http://linkeddatafragments.org/>

Triple Pattern Fragments protocol --

<http://www.hydra-cg.com/spec/latest/triple-pattern-fragments/>

Thank you!