



ARC RDF Store

Easy RDF and SPARQL for LAMP systems

**CSC 8711 Project 4
Topic Report/Manual
Compiled by Akilah McIntyre & EE Durham**

Revision 2

April 22, 2011

Table of Contents:

Table of Contents:.....	2
1.0 History.....	3
2.0 Quick Links.....	4
2.1 ARC2 RDF Store manual links.....	4
3.0 Getting Started	5
3.1 Installation of ARC2	5
3.1.1 Requirements.....	5
3.1.2 Installation of PHP Libraries	5
3.1.3 Installation of MySQL Schema.....	6
3.1.4 Configure the included SPARQL endpoint.....	7
4.0 Using the ARC2 libraries.....	9
4.1 Command line import functionality	9
4.1.1 Loading OWL/RDF Data.....	9
4.1.2 Loading Data via Command Line.....	10
4.1.3 Loading Data via Application Code.....	11
4.2 Querying Against Local Data Using SPARQL	11
4.2.1 Query all triples in the Local RDF Store.....	11
4.2.2 Querying Local Data Using SPARQL:.....	12
4.3 Accessing Data from a Remote Data Store.....	13
4.3.1 Concept of a remote store	13
4.3.2 Querying Remote Data Using SPARQL:	14
5.0 References	18
Articles and Wikis:	18
Data Sites:	19

1.0 History

(courtesy of <https://github.com/semsol/arc2/wiki>)

ARC started in 2004 as a lightweight RDF system for parsing and serializing RDF/XML files. It later evolved into a more complete framework with storage and query functionality. By 2011, ARC2 had become one of the most-installed RDF libraries. Nevertheless, active code development had to be discontinued due to lack of funds and the inability to efficiently implement the ever-growing stack of RDF specifications. The source continues to be available to the community through github.

2.0 Quick Links

2.1 ARC2 RDF Store manual links

1. Getting Started with ARC2: <https://github.com/semsol/arc2/wiki/Getting-started-with-ARC2>
2. Parsing RDF Formats: <https://github.com/semsol/arc2/wiki/Parsing-RDF-Formats>
3. Using the Local Data Store: <https://github.com/semsol/arc2/wiki/Using-ARC%27s-RDF-Store>
4. Using Remote Data Stores: <https://github.com/semsol/arc2/wiki/Remote-Stores-and-Endpoints>
5. Extracting RDF from HTML: <https://github.com/semsol/arc2/wiki/Extracting-RDF-from-HTML>
6. HTTP Reader: <https://github.com/semsol/arc2/wiki/HTTP-Reader>

3.0 Getting Started

3.1 Installation of ARC2

3.1.1 Requirements

- For this demonstration, WAMPServer was used, as available at: <http://www.wampserver.com/en/>
- Basic requirements:
 - a. A web server with PHP5 or PHP4.3 or higher, <http://www.php.net>
 - b. MySQL 5.0 or higher, <http://www.mysql.com>
 - c. Also works in LAMP environment

3.1.2 Installation of PHP Libraries

- 1) In a web browser, go to <https://github.com/tuukka/arc2-starter-pack>
- 2) Download **tuukka-arc2-starter-pack-f9865d2.zip** from <https://github.com/tuukka/arc2-starter-pack/zipball/master>
- 3) Extract **tuukka-arc2-starter-pack-f9865d2** directory to local machine
Result: <local directory>/tuukka-arc2-starter-pack-f9865d2.
- 4) Navigate to the local PHP www root directory, <PHP server>/**www**
Examples:
 - i) C:\Program Files\wamp\www
 - ii) C:\wamp\www
- 5) In **www** directory, create **arc2-starter-pack** subdirectory
Result: www/arc2-starter-pack.
- 6) Copy contents from **tuukka-arc2-starter-pack-f9865d2** directory into **arc2-starter-pack** directory
- 7) The following directory and files should be in the **arc2-starter-pack** directory
 - admin/ <dir>
 - cli.php
 - config.php
 - endpoint.php
 - index.php
 - README
- 8) In **arc2-starter-pack** php directory, create **arc** subdirectory
Result: arc2-starter-pack/arc

- 9) In a web browser, go to <https://github.com/semsol/arc2/>
- 10) Download **semsol-arc2-495d10b.zip** from <https://github.com/semsol/arc2/zipball/master>
- 11) Extract **semsol-arc2-495d10b** directory to local machine
Result: <local directory>/semsol-arc2-495d10b.
Examples:
 - i) C:\Program Files\wamp\www\arc2-starter-pack\arc
 - ii) C:\wamp\www\arc2-starter-pack\arc
- 12) Copy contents from **semsol-arc2-495d10b** directory into **www/arc2-starter-pack/arc** PHP directory
 - a) The following directories and files should be in the **arc2-starter-pack** directory
 - extractors/ <dir>
 - parsers/ <dir>
 - serializers/ <dir>
 - parqlscript/ <dir>
 - store/ <dir>
 - .gitignore
 - ARC2.php
 - ARC2_Class.php
 - ARC2_getFormat.php
 - ARC2_getPreferredFormat.php
 - ARC2_Reader.php
 - ARC2_Resource.php
 - ARC2_TestHandler.php

3.1.3 Installation of MySQL Schema

- 1) In local MySQL server, create database name **arc2test**
create schema arc2test;
- 2) In local MySQL server, create user with all permissions on **arc2test** schema
grant all on arc2test. to 'arc2test_user'@'%' identified by 'RandomPassword';*
- 3) Open config.php file located in the PHP subdirectory arc2-starter-pack and modify credentials for root user local mysql server

Sample Directories:

- i) C:\Program Files\wamp\www\arc2-starter-pack
- ii) C:\wamp\www\arc2-starter-pack

File Modification Example:

```
config.php
// SQL database configuration for storing the postings:
$arc_config = array(
    /* MySQL database settings */
    'db_host' => 'localhost',
    'db_user' => 'arc2test_user',
    'db_pwd' => 'RandomPassword',
    'db_name' => 'arc2test',
```

- 4) To create the tables for the RDF Store, in a web browser, go to <http://localhost:<PHP server port number>/arc2-starter-pack/>

Examples:

- i) <http://localhost:8090/arc2-starter-pack/>
- ii) <http://localhost:8080/arc2-starter-pack/>

Within the index.php file on the default install, the following lines of code will create the RDF Store Tables:

```
$store = ARC2::getStore($config);
if (!$store->isSetUp()) {
    $store->setUp();
}
```

3.1.4 Configure the included SPARQL endpoint

- 1) Edit endpoint.php to point to local sandbox

Sample Directories:

- iii) C:\Program Files\wamp\www\arc2-starter-pack
- iv) C:\wamp\www\arc2-starter-pack

```
$config = array(
    /* db */
    'db_host' => 'localhost',
    'db_name' => 'arc2test',
    'db_user' => 'arc2test_user',
    'db_pwd' => 'RandomPassword',
    /* store name */
    'store_name' => 'sandbox',
```

- 2) To access the SPARQL endpoint, in a web browser, go to <http://localhost:<PHP server port number>/arc2-starter-pack/endpoint.php>

Examples:

- i) <http://localhost:8090/arc2-starter-pack/endpoint.php>
- ii) <http://localhost:8080/arc2-starter-pack/endpoint.php>

4.0 Using the ARC2 libraries

4.1 Command line import functionality

4.1.1 Loading OWL/RDF Data

In the following instructions, we will be using Periodic Table OWL/RDF data from created by Michael Cook, <http://www.daml.org/2003/01/periodictable/PeriodicTable.owl>.

There are two ways to load OWL/RDF data into RDF local store (MySQL database)

- a. Command Line
- b. PHP Application Load

Linux

- o `cd arc2-starter-pack/`
- o `chmod +x cli.php`
- o `./cli.php "LOAD <http://chatlogs.planetrdf.com/swig/2009-07-26>"`
- o `./cli.php "LOAD <file:///home/user/local_file.rdf>"`
- o `./cli.php "LOAD <file://$PWD/file_in_current_dir.ttl>"`
- o `./cli.php "SELECT DISTINCT ?property WHERE { ?subject ?property ?object . }"`
- o `./cli.php "DELETE FROM <http://chatlogs.planetrdf.com/swig/2009-07-26>"`

Windows

- o `cd arc2-starter-pack/`
- o `<PHP Installation>/php.exe /cli.php "LOAD <http://chatlogs.planetrdf.com/swig/2009-07-26>"`
- o `<PHP Installation>/php.exe cli.php "LOAD <file:///home/user/local_file.rdf>"`
- o `<PHP Installation>/php.exe cli.php "LOAD <file://$PWD/file_in_current_dir.ttl>"`
- o `<PHP Installation>/php.exe cli.php "SELECT DISTINCT ?property WHERE { ?subject ?property ?object . }"`
- o `<PHP Installation>/php.exe cli.php "DELETE FROM <http://chatlogs.planetrdf.com/swig/2009-07-26>"`

Sample Windows Output:

Note: these commands are executed from the Windows COMMAND LINE

Note: Make sure that your php executable path is in your systems PATH variable

```
cd C:\Program Files\wamp\www\arc2-starter-pack
```

Sample input

```
php.exe cli.php "LOAD <http://xmlns.com/foaf/spec/index.rdf>"  
Loaded 634 triples.
```

Sample query: Select from Local Data Store

```
php cli.php "SELECT DISTINCT ?property WHERE { ?subject ?property ?object . }"
```

Sample output

```
C:\Program Files\wamp\www\arc2-starter-pack>php cli.php "LOAD  
<http://xmlns.com/foaf/spec/index.rdf>"  
Loaded 634 triples.
```

```
C:\Program Files\wamp\www\arc2-starter-pack>php cli.php "SELECT DISTINCT ?property WHERE {  
?subject ?property ?object . }"  
property  
http://www.w3.org/1999/02/22-rdf-syntax-ns#type  
http://purl.org/dc/elements/1.1/description  
http://purl.org/dc/elements/1.1/title  
http://www.w3.org/2003/06/sw-vocab-status/ns#term_status  
http://www.w3.org/2000/01/rdf-schema#label  
http://www.w3.org/2000/01/rdf-schema#comment  
http://www.w3.org/2000/01/rdf-schema#isDefinedBy  
http://www.w3.org/2000/01/rdf-schema#subClassOf  
http://www.w3.org/2002/07/owl#disjointWith  
http://www.w3.org/2002/07/owl#equivalentClass  
http://www.w3.org/2000/01/rdf-schema#domain  
http://www.w3.org/2000/01/rdf-schema#range  
http://www.w3.org/2000/01/rdf-schema#subPropertyOf  
http://www.w3.org/2002/07/owl#inverseOf  
http://www.w3.org/2002/07/owl#equivalentProperty
```

Sample query: Delete from Local Data Store

```
php cli.php "DELETE FROM <http://xmlns.com/foaf/spec/index.rdf>"
```

4.1.2 Loading Data via Command Line

- 1) In windows environment, add PHP bin to the %PATH% environment
C:/<PHP Server>/bin/php.exe
 - 2) Navigate to the location of **www/arc2-starter-pack** directory
 - 3) Run the following command (all on one line)
> php.exe cli.php "LOAD <http://
http://www.daml.org/2003/01/periodictable/PeriodicTable.owl>"
 - 4) If PeriodicTable.owl was downloaded on to local machine, navigate to the directory containing the data file an run the following command
> php.exe cli.php "LOAD <file:///<location data is
stored>PeriodicTable.owl>"
- If successful, you will see the following output
Loaded 1847 triples.

4.1.3 Loading Data via Application Code

- 1) Navigate to <PHP server >/www/arc2-starter-pack directory.
- 2) Create new php file called, **sample.php**.
- 3) In **load.php**, copy and paste the following lines of code

```
<php?
    include_once("path/to/arc/ARC2.php");

    $config = array(
        /* db */
        'db_name' => 'my_db',
        'db_user' => 'user',
        'db_pwd' => 'secret',
        /* store */
        'store_name' => 'arc_tests',
        /* stop after 100 errors */
        'max_errors' => 100,
    );
    $store = ARC2::getStore($config);
    if (!$store->isSetUp()) {
        $store->setUp();
    }

    $store->query('LOAD
    <http://www.daml.org/2003/01/periodictable/PeriodicTable.owl>');

?>
```

- 4) Open web browser and go to <http://localhost:<PHP server port number>/arc2-starter-pack/load.php>
 - a. Only run this page once or else the data will be inserted into the local RDF store multiple times.

4.2 Querying Against Local Data Using SPARQL

4.2.1 Query all triples in the Local RDF Store

- 1) Navigate to <PHP server >/www/arc2-starter-pack directory.
- 2) Create new php file called, **local_query_triple.php**.
- 3) Copy and paste the following lines of code in **local_query_triple.php** in to execute a SPARQL Query to list each subject, object, domain triple in RDF store.

```
<?php
    include_once("arc/ARC2.php");
    include_once('config.php');

    $store = ARC2::getStore($arc_config);

    if (!$store->isSetUp()) {
        $store->setUp(); /* create MySQL tables */
    }

    $q = '
        SELECT DISTINCT ?subject ?property ?object WHERE {
```

```
        ?subject ?property ?object .
    }
';

$rows = $store->query($q, 'rows');

$r = '';
if ($rows = $store->query($q, 'rows')) {
    $r = '<table border=1>
        <th>Subject</th><th>Property</th><th>Object</th>'. "\n";
    foreach ($rows as $row) {
        $r .= '<tr><td>'. $row['subject'] .
            '</td><td>'. $row['property'] .
            '</td><td>'. $row['object'] . '</td></tr>'. "\n";
    }
    $r .= '</table>'. "\n";
}
else{
    $r = '<em>No data returned</em>';
}

echo $r;
?>
```

- 4) Open web browser and go to http://localhost:<PHP server port number>/arc2-starter-pack/local_query_triple.php

4.2.2 Querying Local Data Using SPARQL:

Using query from <http://www.xml.com/pub/a/2005/11/16/introducing-sparql-querying-semantic-web-tutorial.html?page=4>, find the atomic weights and CAS registry numbers of halogens and noble gases.

- 1) Navigate to <PHP server >/www/arc2-starter-pack directory.
- 2) Create new php file called, **local_query_periodic.php**.
- 3) Copy and paste the following lines of code in **local_query_periodic.php**.

```
<?php
include_once("arc/ARC2.php");
include_once('config.php');

    $store = ARC2::getStore($arc_config);

    if (!$store->isSetUp()) {
        $store->setUp(); /* create MySQL tables */
    }

    $q = 'PREFIX table:
<http://www.daml.org/2003/01/periodictable/PeriodicTable#>
SELECT *
FROM <http://www.daml.org/2003/01/periodictable/PeriodicTable.owl>
WHERE
```

```
{
  {
    ?element table:name ?name;
          table:symbol ?symbol;
          table:atomicNumber ?number;
          table:group table:group_17.
  }
  UNION
  {
    ?element table:name ?name;
          table:symbol ?symbol;
          table:atomicNumber ?number;
          table:group table:group_18.
  }
};

$rows = $store->query($q, 'rows');

$r = '';
if ($rows = $store->query($q, 'rows')) {
  $r = '<table border=1>
<th>Name</th><th>Symbol</th><th>Number</th>'. "\n";
  foreach ($rows as $row) {
    $r .= '<tr><td>'. $row['name'] .
      '</td><td>'. $row['symbol'] .
      '</td><td>'. $row['number'] . '</td></tr>'. "\n";
  }
  $r .= '</table>'. "\n";
}
else{
  $r = '<em>No data returned</em>';
}

echo $r;
?>
```

- 4) Open web browser and go to http://localhost:<PHP server port number>/arc2-starter-pack/local_query_periodic.php

4.3 Accessing Data from a Remote Data Store

4.3.1 Concept of a remote store

So one can load remote data into a local data store and run fast, local queries. But the problem remains that data changes. And keeping up the changes is the natural purview of the data owner, not an application developer.

ARC2 addresses this problem with the concept of the Remote Data Store. ARC2 allows the store initialization clause to address a remote data store instead of a local data store, and still access the data in the same exact fashion that one would for local data.

Documentation Link: <https://github.com/semsol/arc2/wiki/Remote-Stores-and-Endpoints>

Configuration format :

```
/* ARC2 static class inclusion */
include_once('path/to/arc/ARC2.php');

/* configuration */
$config = array(
    /* remote endpoint */
    'remote_store_endpoint' => 'http://example.com/sparql',
);

/* instantiation */
$store = ARC2::getRemoteStore($config);
```

4.3.2 Querying Remote Data Using SPARQL:

Example 1

Using the remote data store from the Linked Movie Database (linkedmdb.org), find actors who've worked with Kevin Bacon and the number of movies in which they've worked together.

- 1) Navigate to <PHP server >/www/arc2-starter-pack directory.
- 2) Create new php file called, **remote_query_movies.php**.

Copy and paste the following lines of code into the new file, **remote_query_movies.php**:

```
<?php
include("config.php");

$remote_store_endpoint = 'http://data.linkedmdb.org/sparql';

/* configuration */
$config = array(
    /* remote endpoint (gene database)*/
    'remote_store_endpoint' => $remote_store_endpoint,
);

/* instantiation */
$store = ARC2::getRemoteStore($config);

$q = '
SELECT DISTINCT (COUNT(?kb) AS ?movieCount) ?actorName WHERE {
    ?kb    <http://data.linkedmdb.org/resource/movie/actor_name> "Kevin Bacon".
    ?movie <http://data.linkedmdb.org/resource/movie/actor> ?kb;
          <http://data.linkedmdb.org/resource/movie/actor> ?actor.
    ?actor <http://data.linkedmdb.org/resource/movie/actor_name> ?actorName.
    FILTER (?kb != ?actor).
}GROUP BY ?actorName ORDER BY ?actorName
';
```

```
$rows = $store->query($q, 'rows');

$result = $store->query("");

print "<HTML>\n";
print "<HEAD>\n";
print "<TITLE>Actors who've been in a movie with Kevin Bacon</TITLE>\n";
print "<link rel='stylesheet' type='text/css'
href='http://data.linkedmdb.org/snorql/style.css' /> \n";

print "</HEAD>\n";
print "<BODY>\n";
print "<h1>Actors who've been in a movie with Kevin Bacon</h1>\n";
print "<p>\n";
print "</p>\n";
print "<TABLE BORDER=1 class='queryresults'>\n";
print "<TR>\n";
print "<TH>Movie Count</TH>\n";
print "<TH>Actor</TH>\n";
print "</TR>\n";
$row_counter = 0;
if ($rows = $store->query($q, 'rows')) {
    foreach ($rows as $row) {
        print "<TR ";
        if ($row_counter%2 == 0)
        {
            print "class='even'";
        } else
        {
            print "class='odd'";
        }
        print " >\n";
        print "<TD> " . $row['movieCount'] . " </TD><TD>" . $row['actorName'] .
"</TD>\n";
        print "</TR>\n";
        $row_counter ++;
    }
} //end if --non-empty results
else
{
    print "<TR>\n";
    print "<TD>no data found</TD>\n";
    print "</TR>\n";
} //end else --empty results
print "</TABLE>\n";
print "<p>\n";
print "</p>\n";

print "</BODY></HEAD>\n";

?>
```

Example 2

Using the remote data store from the Gene Ontology Database (www.obofoundry.org), find cellular processes that are either integral to, or a refinement of, signal transduction.

- 1) Navigate to <PHP server >/www/arc2-starter-pack directory.
- 2) Create new php file called, **remote_query_gene.php**.

Copy and paste the following lines of code in **remote_query_gene.php**.

```
<?php
include("config.php");

$remote_store_endpoint = 'http://data.linkedmdb.org/sparql';

/* configuration */
$config = array(
    /* remote endpoint (gene database)*/
    'remote_store_endpoint' => $remote_store_endpoint,
);

/* instantiation */
$store = ARC2::getRemoteStore($config);

$q = '
SELECT DISTINCT (COUNT(?kb) AS ?movieCount) ?actorName WHERE {
    ?kb    <http://data.linkedmdb.org/resource/movie/actor_name> "Kevin Bacon".
    ?movie <http://data.linkedmdb.org/resource/movie/actor> ?kb;
           <http://data.linkedmdb.org/resource/movie/actor> ?actor.
    ?actor <http://data.linkedmdb.org/resource/movie/actor_name> ?actorName.
    FILTER (?kb != ?actor).
}GROUP BY ?actorName ORDER BY ?actorName
';

$rows = $store->query($q, 'rows');

$result = $store->query("");

print "<HTML>\n";
print "<HEAD>\n";
print "<TITLE>Actors who've been in a movie with Kevin Bacon</TITLE>\n";
print "<link rel='stylesheet' type='text/css'
href='http://data.linkedmdb.org/snorql/style.css' /> \n";

print "</HEAD>\n";
print "<BODY>\n";
print "<h1>Actors who've been in a movie with Kevin Bacon</h1>\n";
print "<p>\n";
print "</p>\n";
print "<TABLE BORDER=1 class='queryresults'>\n";
print "<TR>\n";
print "<TH>Movie Count</TH>\n";
```



```
print "<TH>Actor</TH>\n";
print "</TR>\n";
$row_counter = 0;
if ($rows = $store->query($q, 'rows')) {
    foreach ($rows as $row) {
        print "<TR ";
        if ($row_counter%2 == 0)
        {
            print "class='even'";
        } else
        {
            print "class='odd'";
        }
        print " >\n";
        print "<TD> " . $row['movieCount'] . " </TD><TD>" . $row['actorName'] .
"</TD>\n";
        print "</TR>\n";
        $row_counter ++;
    }
} //end if --non-empty results
else
{
    print "<TR>\n";
    print "<TD>no data found</TD>\n";
    print "</TR>\n";
} //end else --empty results
print "</TABLE>\n";
print "<p>\n";
print "</p>\n";

print "</BODY></HEAD>\n";

?>
```

5.0 References

Articles and Wikis:

1. arc-dev Google Group
<http://groups.google.com/group/arc-dev>
2. ARC2, 2010. Hendler, McGuinness, Ding et al.
<http://data-gov.tw.rpi.edu/wiki/ARC2>
3. ARC2 Wiki, 2011. Nowack, Benjamin.
<https://github.com/semsol/arc2/wiki/>
4. Linked Data, 2009. Heath, Tom.
<http://linkeddata.org/>
5. The Linking Open Data cloud diagram, 2010. Cyganiak, Richard.
<http://richard.cyganiak.de/2007/10/lod/>
6. RDF SPARQL Endpoint, 2011. Corlosquet, Stéphane.
http://drupal.org/project/sparql_ep
7. SPARQL and RDF stores for SMW, 2010. Krötzsch, Markus.
http://semantic-mediawiki.org/wiki/SPARQL_and_RDF_stores_for_SMW
8. Semantic Web SPARQL end-points, 2010. Burleson, Cody.
<https://wiki.base22.com/display/btg/Semantic+Web+SPARQL+end-points>
9. SPARQL, 2011. Prud'hommeaux, Eric Gordon .
<http://en.wikipedia.org/wiki/SPARQL>
10. SPARQL Query Language for RDF, 2008. W3C.
<http://www.w3.org/TR/rdf-sparql-query/>
11. SparqlEndpoints, 2009. W3C.
<http://www.w3.org/wiki/SparqlEndpoints>
12. Using PHP in Linked Data Applications, 2010. Bao, Jie.
<http://dig.csail.mit.edu/2010/Courses/6.898/SWLibraries/php/ARC2>
13. WE WANT RAW DATA NOW, 2009. Berners-Lee, Tim.
http://www.ted.com/talks/tim_berniers_lee_on_the_next_web.html

Data Sites:

1. Linked Movie Database
<http://www.linkedmdb.org/>
2. The Open Biological and Biomedical Ontologies
<http://www.obofoundry.org/>
3. UK Government datasets
<http://data.gov.uk/data>
4. U.S. Federal Executive Branch datasets
<http://www.data.gov/catalog/raw>
5. U.S. Office of Personnel Management
<http://www.fedscope.opm.gov/>