

Alliance of Information and Referral Systems (AIRS)

Linked Data Pilot Project

*Final Report March 2016*

|  |  |
| --- | --- |
|  |  |

Table of Contents

[Executive Summary](#h.6nmbb416v5sd)

[Background](#h.e493416llgos)

[Demonstration](#h.7wdleez4r5tj)

[Phase 1](#h.z3fe36u2in6d)

[A New AIRS Web Vocabulary](#h.bewyt736fhtc)

[Eight Site Pilot Demonstrating Use The New AIRS Web Vocabulary](#h.3a645nt4ookm)

[Phase 2](#h.i09pafkha914)

[Overall Gateway Demonstration Methodology](#h.l6iu589cf5l4)

[Gateway Implementation Technical Overview](#h.b9mue3izkmgd)

[Gateway Demonstration Artifacts And Screenshots](#h.7asew2b5xvap)

[Design](#h.cbudtryup7xg)

[Development](#h.b0ql0idncdyp)

[Image: Configuring the controlling web server](#h.d8jg0gtg6xvf)

[Image: Setting up a developer account with the API Gateway service provider.](#h.nln5eknnymfc)

[Image: Setting up the APIs that the API Gateway service provider.](#h.rlbpxrcevghi)

[Image: Configuring what your data customers are allowed to do.](#h.g02s5kqzsdfv)

[Image: Setting up payment plans.](#h.r4vxke1feh5w)

[Deployment](#h.mucd09o0beox)

[Image: Mock 3rd party web app running in browser.](#h.20dz9dothjaj)

[Image: Retrieving the protected linked data after supplying the correct access token.](#h.nd9svw6rpdp)

[Testing](#h.xzecy0futiwc)

[Images: Various images of sending deconstructed test messages.](#h.bpke9e48qd3b)

[Example Analytics On Access To The AIRS Linked Data](#h.7vk6sfhnqgsu)

[Image: See total usage. You can also see usage by app.](#h.w4v5c2sjuumq)

[Image: See what hours of the day your customers are accessing data.](#h.15dyt2rab1mj)

[Image: Daily data usage/traffic monitoring.](#h.iwtp21hrss4a)

[Appendix 1 - Listing of Pilot Sites](#h.8a1z5hi66z24)

[Appendix 2 - Screenshot Of Example Phase 1 Multi-location Service Lookup](#h.f3kqtmaq470g)

[Appendix 3 - 3rd Party Application Data Access Workflow, using OAuth 2](#h.n7peqn5cqlmn)

# 

# 

# Executive Summary

Linked Data is the most appropriate technology for providing web searchable AIRS resource data. This document details the successful pilot demonstration of using Linked Data to share and search for information and referral data across many sites and regions. It then shows how we layered optional security, payment, other terms, and usage restrictions upon the shared Linked Data. This is mainstream technology ready for deployment by AIRS compliant information and referral service system vendors and service providers.

This AIRS project is open source and available for all AIRS members, partners, vendors, collaborators or outside entities, to take to the next level, under a "business friendly" license[[1]](#footnote-1) (no restrictions). AIRS can provide referrals to technical implementation resources.

An implementation stage would require the development of a local action plan which should include:

* resource specialist and IT support training on the basics of W3C Linked Data;
* the cooperation of the existing resource database system vendor, possibly facilitated by a third-party, in hosting the AIRS Linked Open Vocabulary;
* investment in technical expertise, potentially through software vendor, to undertake detailed implementation and project management oversight; and
* security arrangements (likely through selection of a gateway vendor like 3scale or the Amazon Gateway).

# Background

In an effort to explore improved information and referral interoperability, the Alliance of Information and Referral Systems (AIRS) Board commissioned the discovery of appropriate technologies and protocols to enable these improvements. A preliminary Board report[[2]](#footnote-2) was released with recommendations for these technologies and protocols, followed by Board approval of a two phased demonstration of the technologies. To summarize, the technology recommended was W3C Linked Data[[3]](#footnote-3), originally introduced by Tim Berners-Lee, creator of the World Wide Web. It reuses commonly existing web technologies, to take the Web into the next level of coordination and interoperability. Linked Data is superior to standard web APIs for exposing libraries of data with their underlying relationships intact. Information and Referral organizations typically maintain a library of provider agencies and services available at specific locations, making it a good fit with Linked Data.

The purpose of this follow-up report is to describe the activities and outcomes of the two demonstration phases, and an explanation of how each phase was able to successfully achieve its goal and demonstrate a path forward toward fully controllable, real-time resource information sharing between different software provider systems. The first phase has already been reported to the AIRS Board previously, so the second phase is given more detailed treatment within this document. Sharing client referral information is outside the scope of this demonstration.

Special thanks to the AIRS Resource Specialists who assisted in the development of the AIRS Linked Open Vocabulary, to the eight pilot site participants, and to the AIRS Board for nurturing this technology application to improve information and referral interoperability.

# Demonstration

The demonstration’s overall purpose across both phases is to show the feasibility of using Linked Data *within the context of AIRS compliant information and referral resource sharing.* Linked Data is already commonly used in library and public information repository contexts, so we simply wanted to demonstrate the use of the Linked Data to share AIRS defined resource information on agencies, sites, and services.

## Phase 1

The goal of Phase 1 was to show that data elements found within the existing AIRS Glossary[[4]](#footnote-4) could be translated into a standard web vocabulary[[5]](#footnote-5). Examples of these data elements are “Agency”, “Site”, and “Service”. This vocabulary step was necessary for standardizing how AIRS Glossary concepts are related and represented as data; an extra level of specificity required for software to interoperate. Once we had created the web vocabulary, we wanted to show how this new vocabulary could be used by diverse AIRS member organizations to share their sample data in real time.

### A New AIRS Web Vocabulary

To achieve this, we started by creating the web vocabulary with the help of volunteer resource specialists recruited on the AIRS Networker forums. The discussion and process of arriving at the appropriate relationships between AIRS Glossary concepts and representing them in a complete AIRS Linked Open Vocabulary (LOV) is documented in an issue tracker,[[6]](#footnote-6) and also in an archived AIRS Networker online forum topic on the subject.[[7]](#footnote-7) The working group started with the data elements encountered within the AIRS XML Schema v3.1[[8]](#footnote-8), and added previously and newly identified enhancements, itemized within the issue tracker. The first complete version of the web vocabulary version was completed in June 2014, with five subsequent updates committed through February 2015.[[9]](#footnote-9)

### Eight Site Pilot Demonstrating Use The New AIRS Web Vocabulary

With the web vocabulary ready for use, we located an initial eight pilot[[10]](#footnote-10) sites (see listing of sites in [Appendix 1](#h.8a1z5hi66z24)) to represent their information as AIRS Linked Open Data, following the conventions within the newly created web vocabulary. The eight pilot sites represented a diverse selection of information and referral organization types, distributed geographically throughout the United States and Canada.

First, in early 2015, the pilots extracted a minimum viable sample set of data fields from their live, production data sets.[[11]](#footnote-11) Then, each pilot site converted their respective sample data set to the AIRS Linked Open Vocabulary syntax, making it into valid W3C Linked Data. In May 2015, we then loaded each site’s Linked Data onto a separate server on The Cloud, to simulate separate information and referral sites sharing and referencing each other’s Linked Data.[[12]](#footnote-12) Unlike Phase 2 of the demonstration to come, all data was posted publicly, without access restrictions. We simply wanted to show that the sharing and look-up capability was functional. See [Appendix 2](#h.uirykj1frynj) for a demonstration screenshot of multi-site information look-up. In June 2015, we held a contest for querying each other sites’ data over the web. Phase 1 successfully achieved its goals of demonstrating unrestricted sharing of Linked Data amongst a diverse set of Information and Referral organizations. Next, we needed to demonstrate a single pilot site’s ability to flexibly and dynamically restrict access to its Linked Data.

## Phase 2

Now that unfettered AIRS data sharing was demonstrated in Phase 1, Phase 2 would show that it was possible to do the opposite: restrict access dynamically to a Linked Data set published by a pilot site. Information and Referral organizations often need to protect access to their library of services, since the collected data may be viewed as copyrighted property critical to their business model. By exposing their data over the web as Linked Data, access to the data can easily be conditionally granted to 3rd party data consumers, by the use of a secure Application Programming Interface (API)[[13]](#footnote-13) Gateway. Conditions to 3rd party data access, enabled by a Gateway, may include:

* payment for access to the data,
* throttling excessive use of the data, and
* acceptance of terms of use (deletion of data after a period of time, prohibiting further sharing of the obtained data, etc.).

### Overall Gateway Demonstration Methodology

The approach taken to implement protected AIRS Linked Data was to copy one of the eight pilot sites to a new server, and then add a Gateway in front of it. We selected an online Gateway service provider, instead of building our own from scratch, since many alternative low cost vendors were available to select from.[[14]](#footnote-14) We chose the service of a company named [3scale.net](http://www.3scale.net/).[[15]](#footnote-15) So, once the 3scale Gateway service was deployed, then access control, usage analytics, and preconditions to 3rd party data access were configured. Then, we created a fictitious web app, representing a 3rd party data customer, seeking access to the site’s protected data, and tested obtaining access to the data. As for the authentication protocol used, we chose OAuth 2[[16]](#footnote-16), which is a very popular technology used to provide secure access to contemporary web application frameworks.

### Gateway Implementation Technical Overview

First, we had to set up a new web server that would control access to the protected Linked Data (see [Appendix 3](#h.n7peqn5cqlmn), step 2, the box labeled “nginx”). All communication to the protected Linked Data set would now had to pass through this web server, which blocks access by default. We installed a series of open source control scripts into this web server[[17]](#footnote-17), which 3scale provided us. The scripts control the specific way that communication with the 3rd party data requestor is orchestrated, and various other options are offered by 3scale, depending on the specific situation. We chose the “Client Credentials Flow”[[18]](#footnote-18) Then, we informed the Gateway service provider of our web server details, so 3scale would expect access requests from this controlling web server. So, in a nutshell, if a 3rd party data requestor wants the AIRS Linked Data, they have to get through the web server, which in turn requires the approval of 3scale.net. So, the 3rd party data requestor must first select an AIRS Linked Data data plan at 3scale’s website. When the 3rd party data requestor makes its request for the data, 3scale already knows what plan they have selected and perhaps paid for, and 3scale instructs our web server to allow them access to the AIRS Linked Data, under whatever conditions the plan enforces (i.e. data caps).

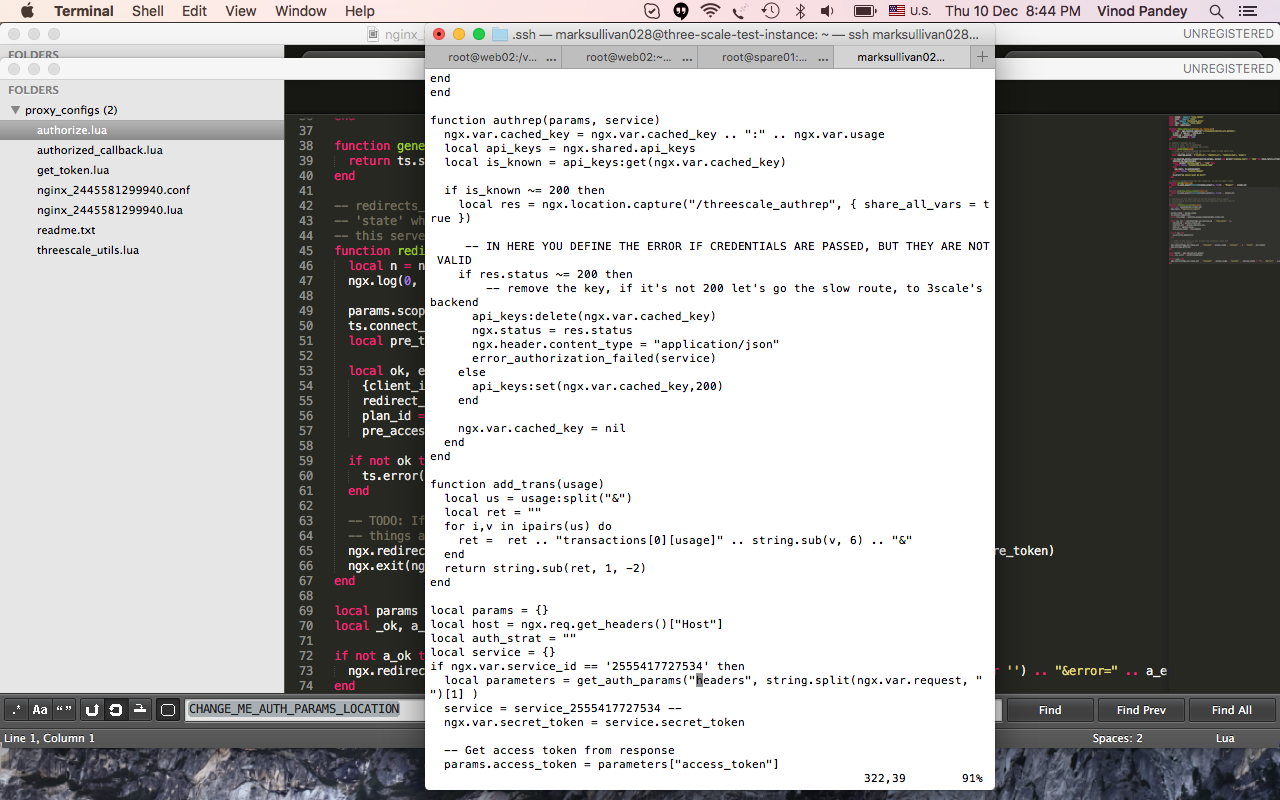
### Gateway Demonstration Artifacts And Screenshots

The code for the Gateway demonstration is at: <https://github.com/airsalliance/gateway_implementation_demo>. It is freely licensed open source code anyone may use. The code consists of the controlling web server configuration scripts, the mock 3rd party data requesting web site (playing the role of an I&R software vendor accessing data from other parts of the country/world), as well as modifications to the Ontario Linked Data server so that it would only accept requests via the controlling web server.

#### Design

We only implemented existing protocols. See [Appendix 3](#h.n7peqn5cqlmn) for workflow and documentation on the “Client Credentials Flow”[[19]](#footnote-19) of the OAuth 2 protocol.

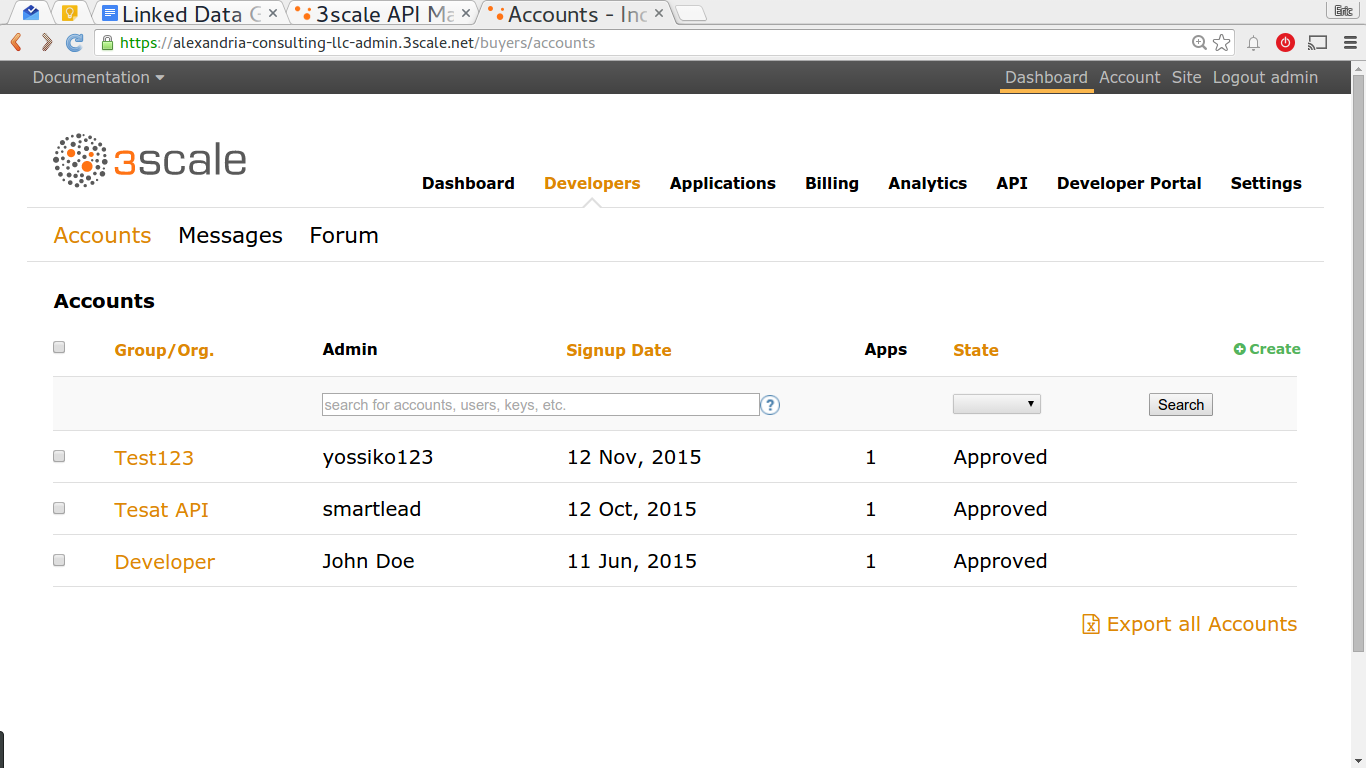
#### Development



##### Image: Configuring the controlling web server

The web server (specifically an nginx proxy server) that communicates with the API Gateway service provider (in this case, 3Scale.net).

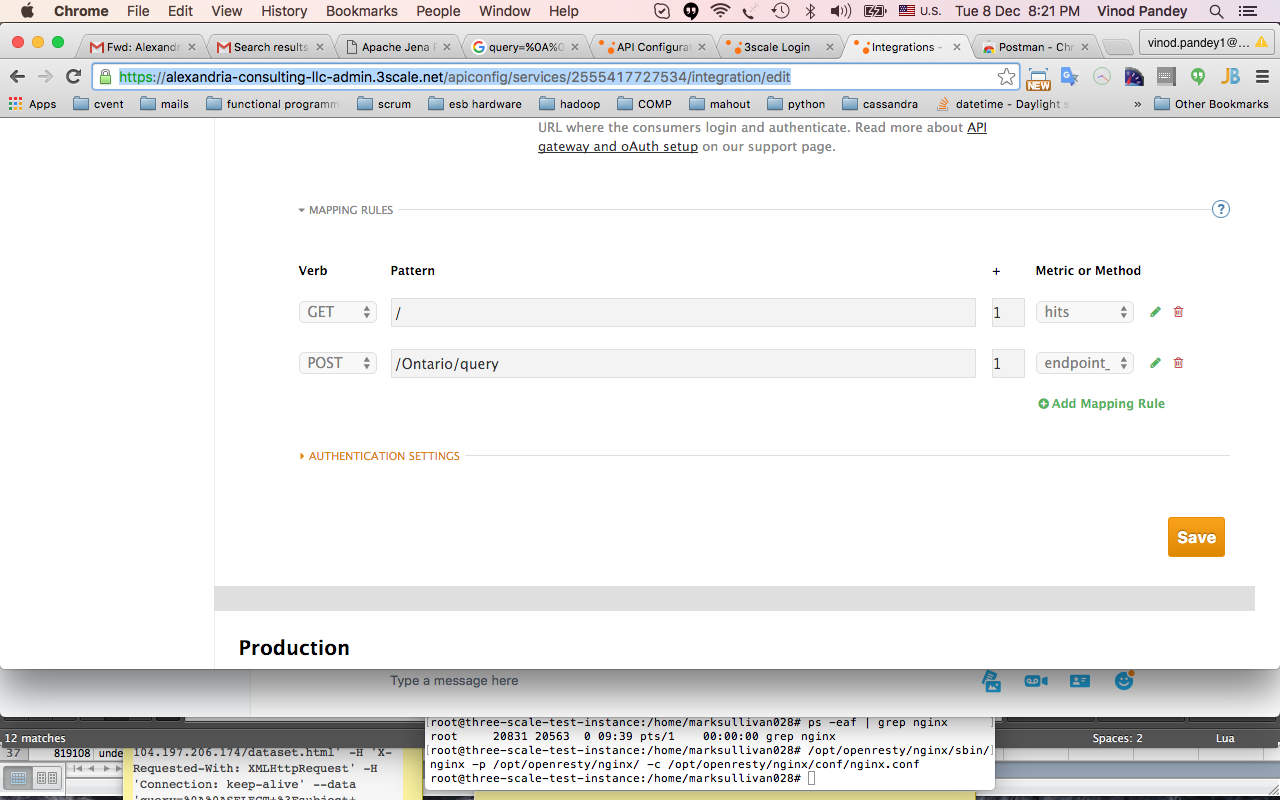
##### 



##### Image: Setting up a developer account with the API Gateway service provider.

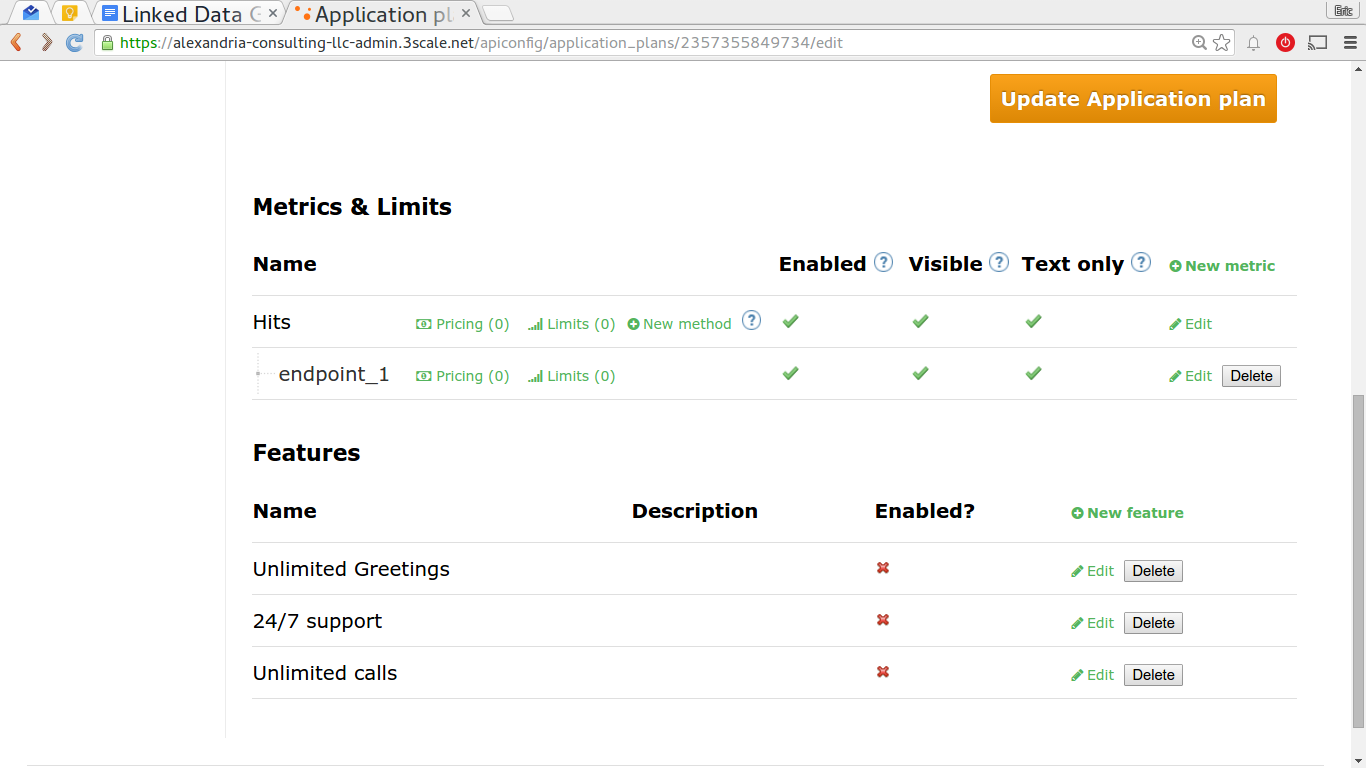
Each I&R vendor would have to establish a similar relationship with their selected API Gateway service provider (in this case, 3scale.net), if the vendor chooses to outsource this functionality.

##### 

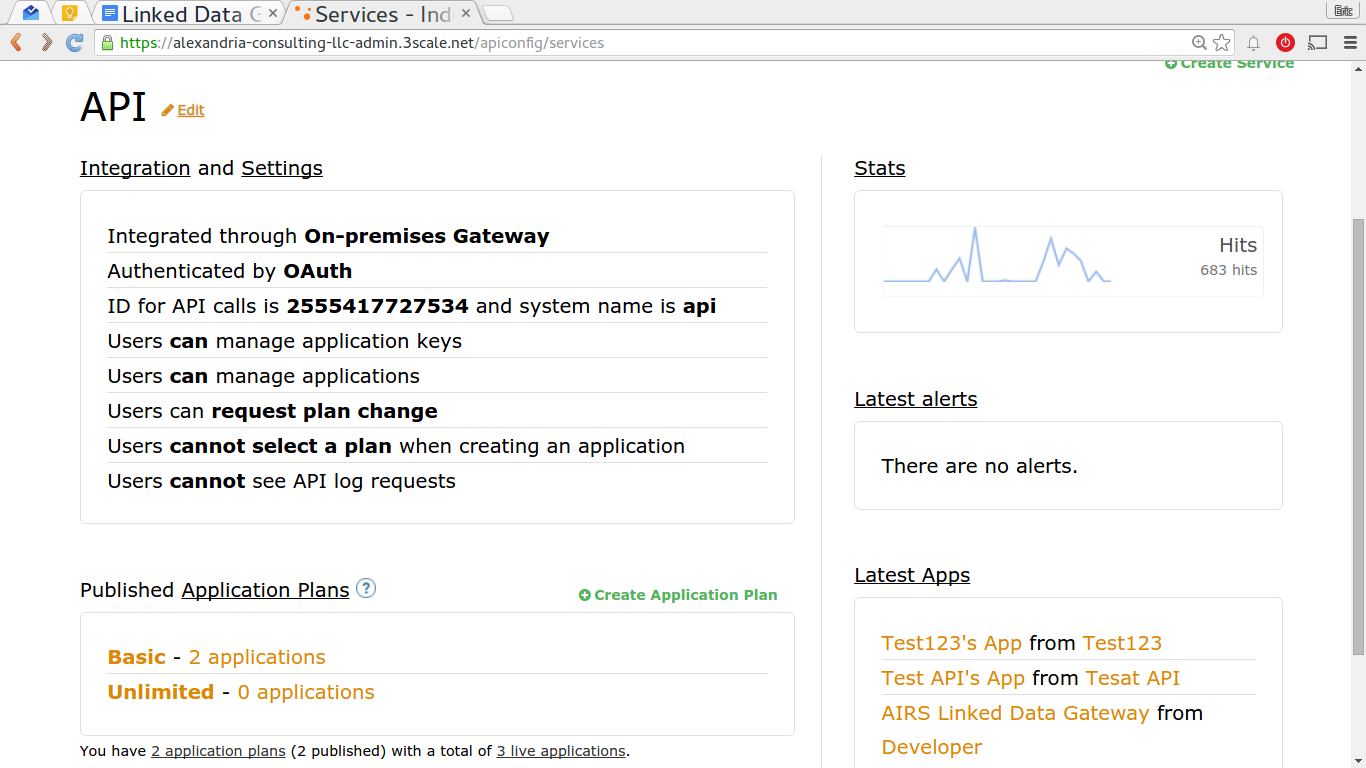


##### Image: Setting up the APIs that the API Gateway service provider.

In this case, the API Gateway service provider is 3scale.net, which must be made aware of which APIs it must control access to.

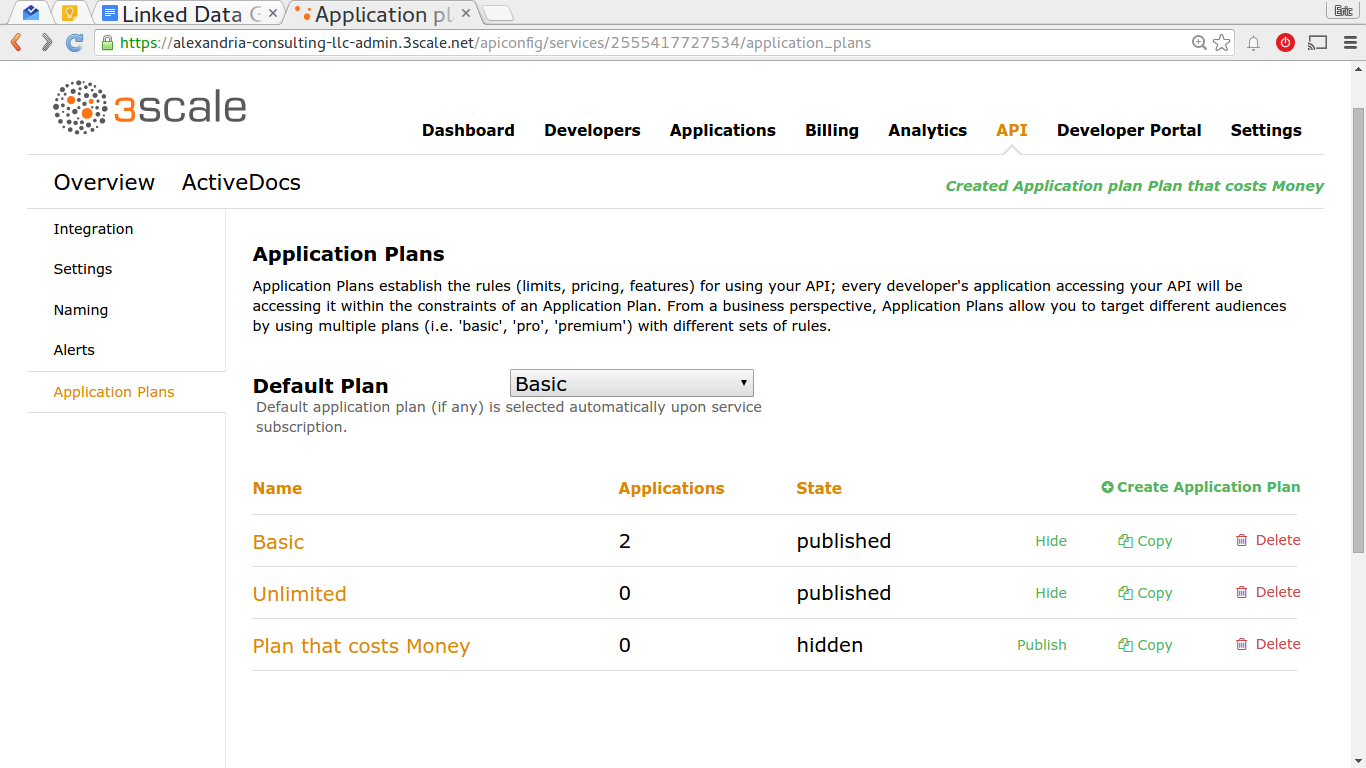


*Image: Throttling data usage.*



##### Image: Configuring what your data customers are allowed to do.

##### 

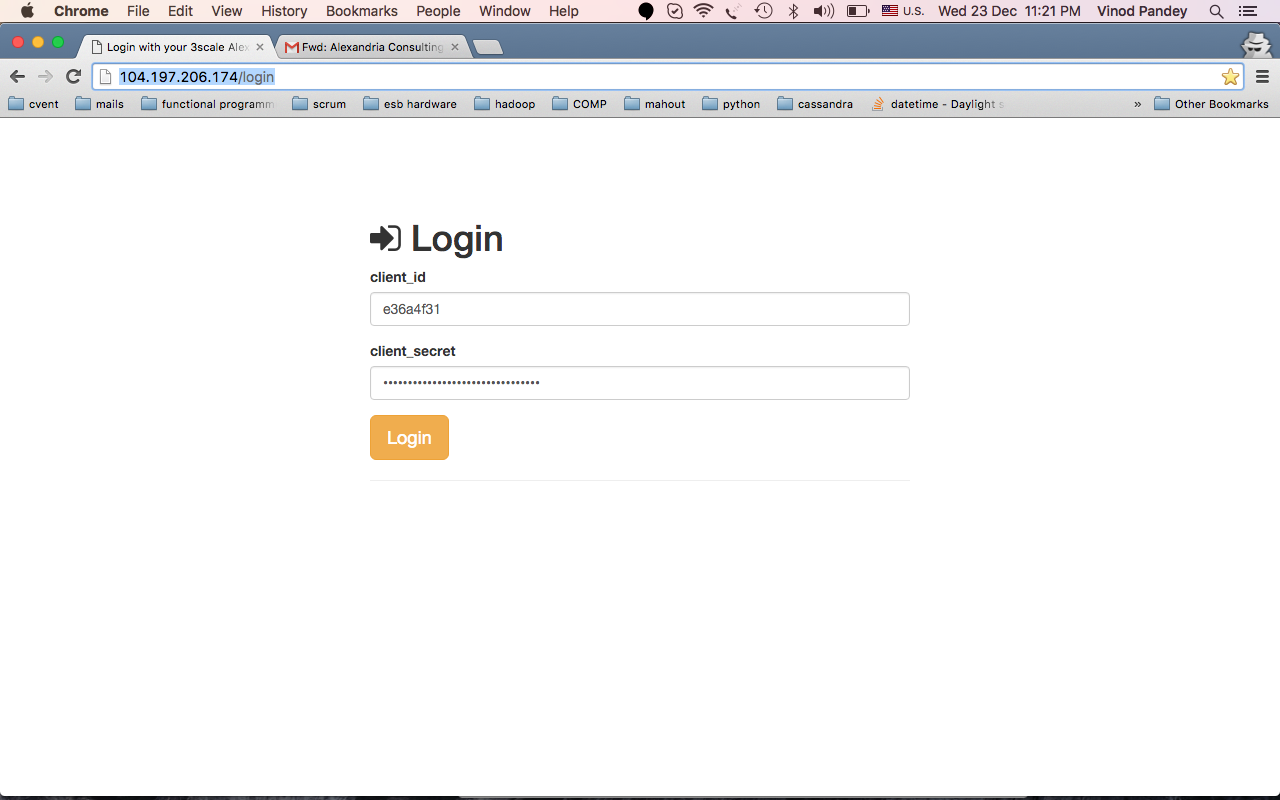


##### Image: Setting up payment plans.

An API Gateway typically allows you to charge money for access to your secured data, if your agency desires.

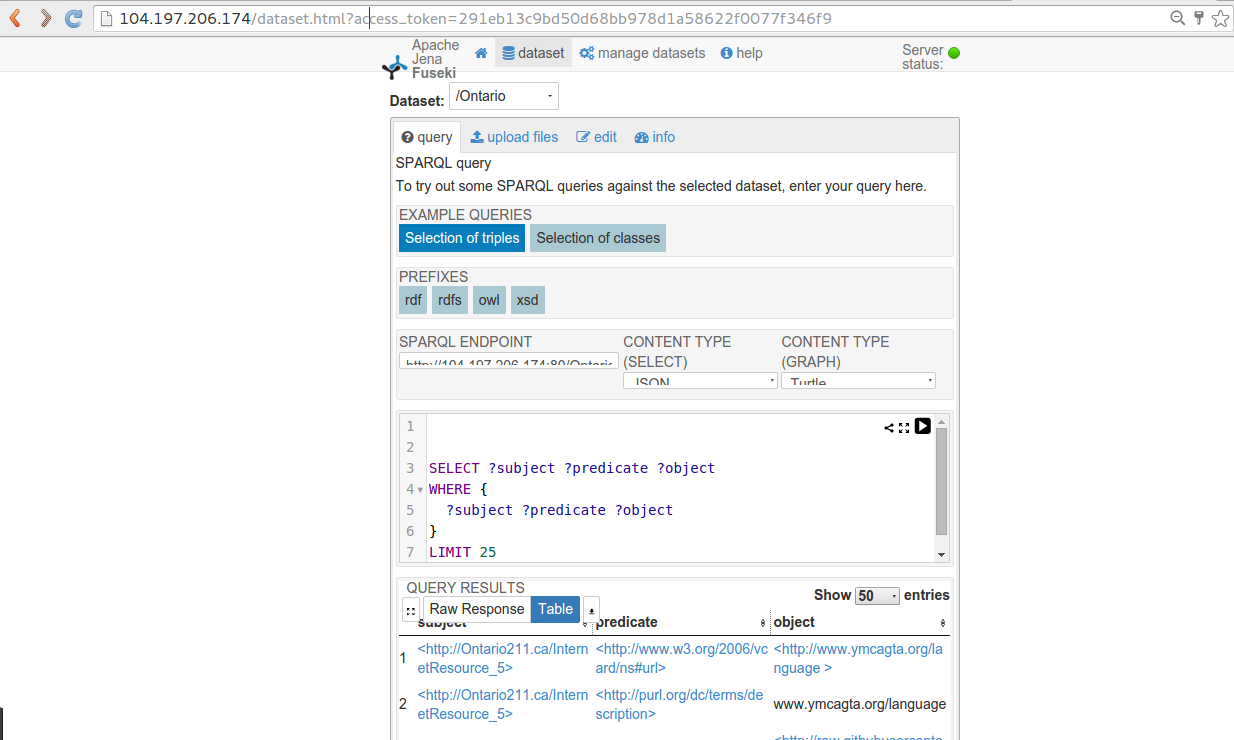
#### 

#### Deployment



##### Image: Mock 3rd party web app running in browser.

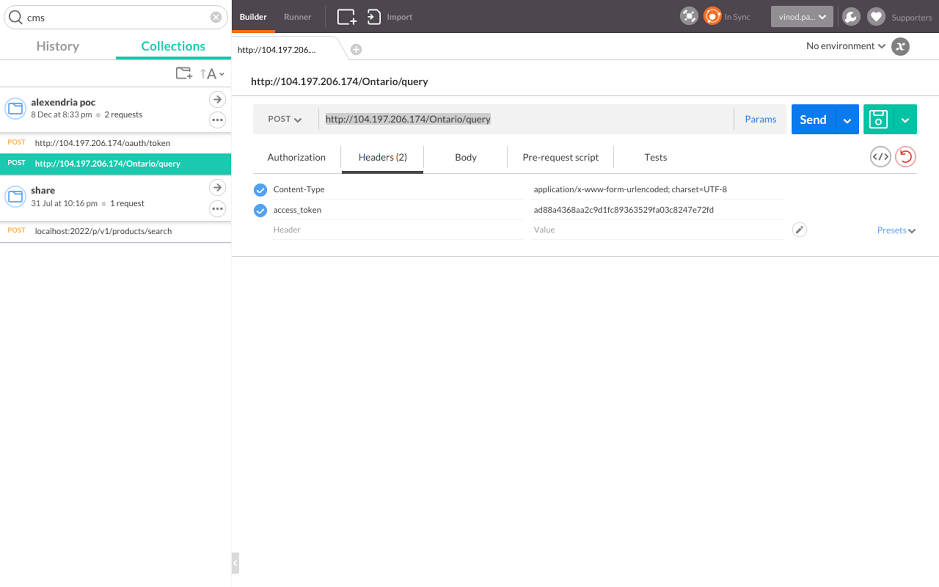
If this mock app supplies the right credentials that 3Scale considers a valid/paid account, it can access the Ontario AIRS Linked Data. The actual url for this login page is <http://104.197.206.174/login>.[[20]](#footnote-20)

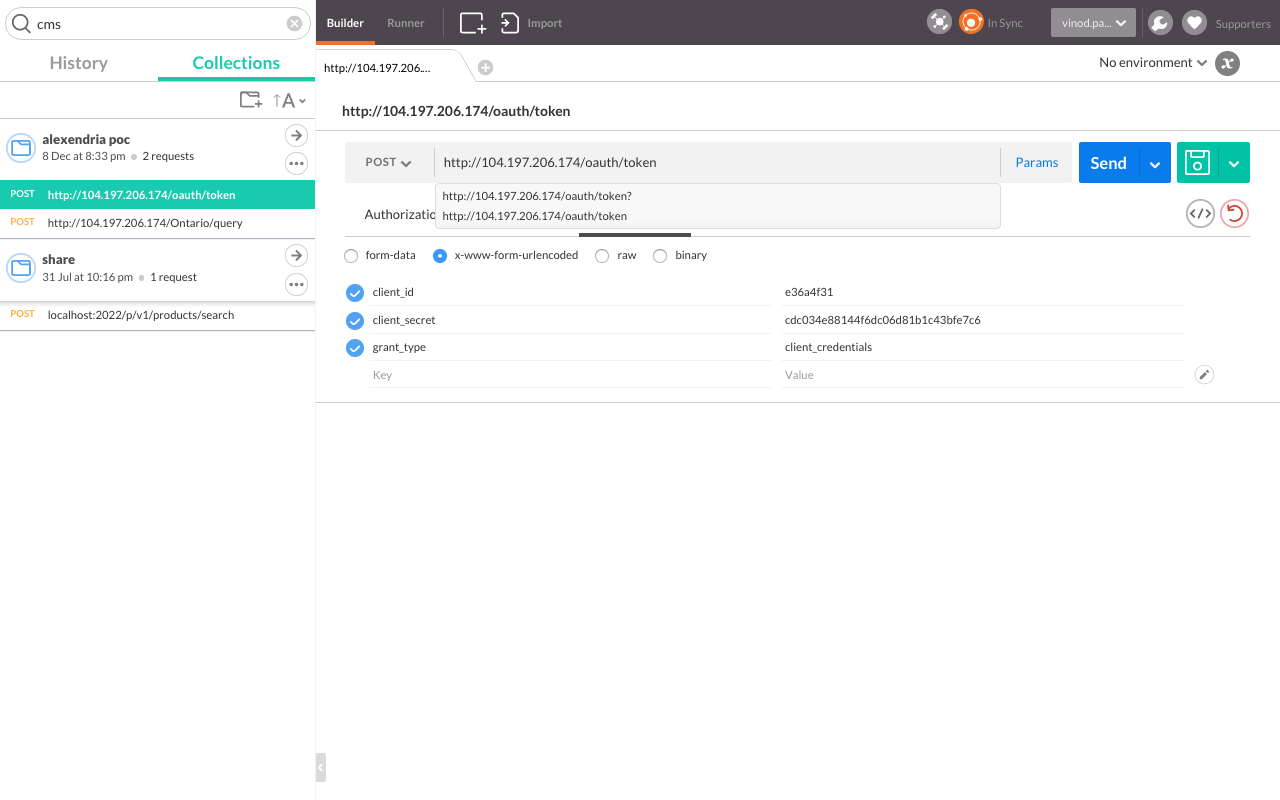


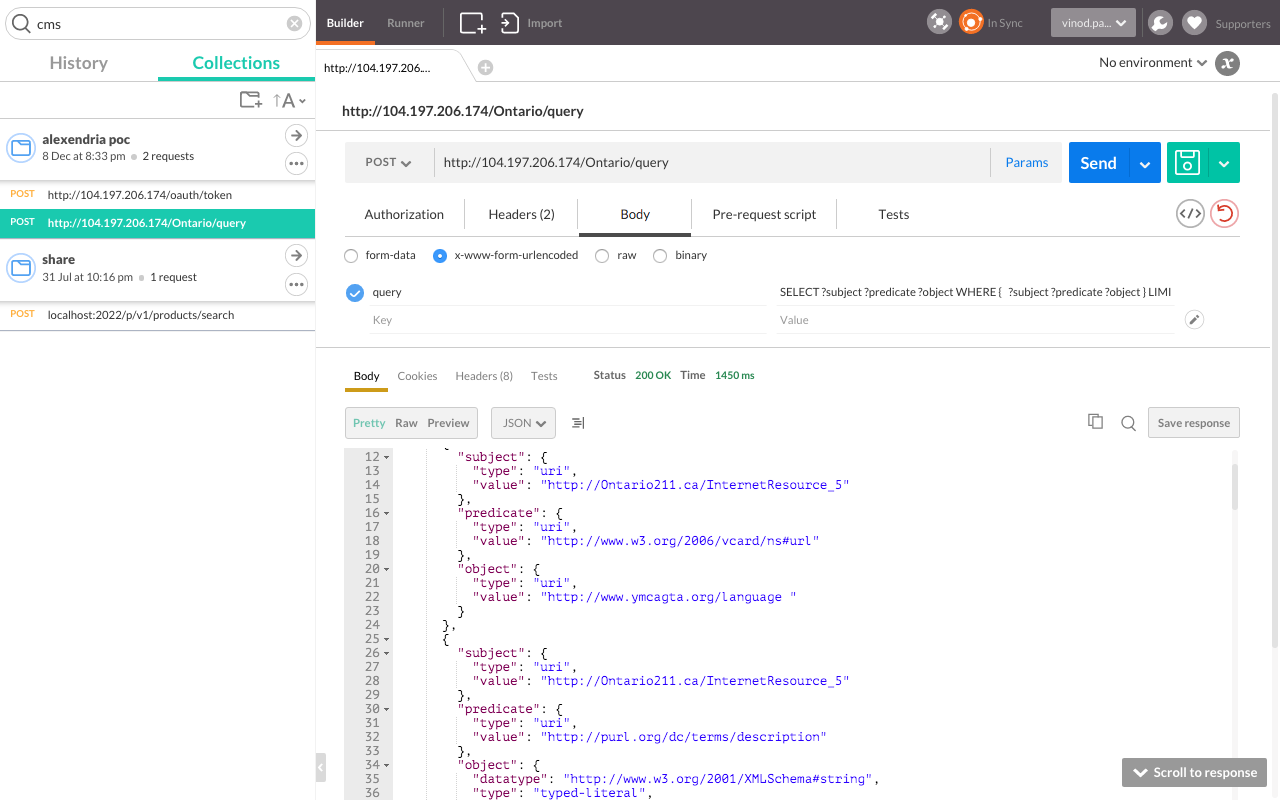
##### Image: Retrieving the protected linked data after supplying the correct access token.

This screen appears after the previous mock app login screenshot. We embedded a query tool in the mock web app to show the data we got back from the protected Ontario Linked Data Pilot Site (notice the access token passed in the url; it needs to supply this when it requests the data).

#### Testing







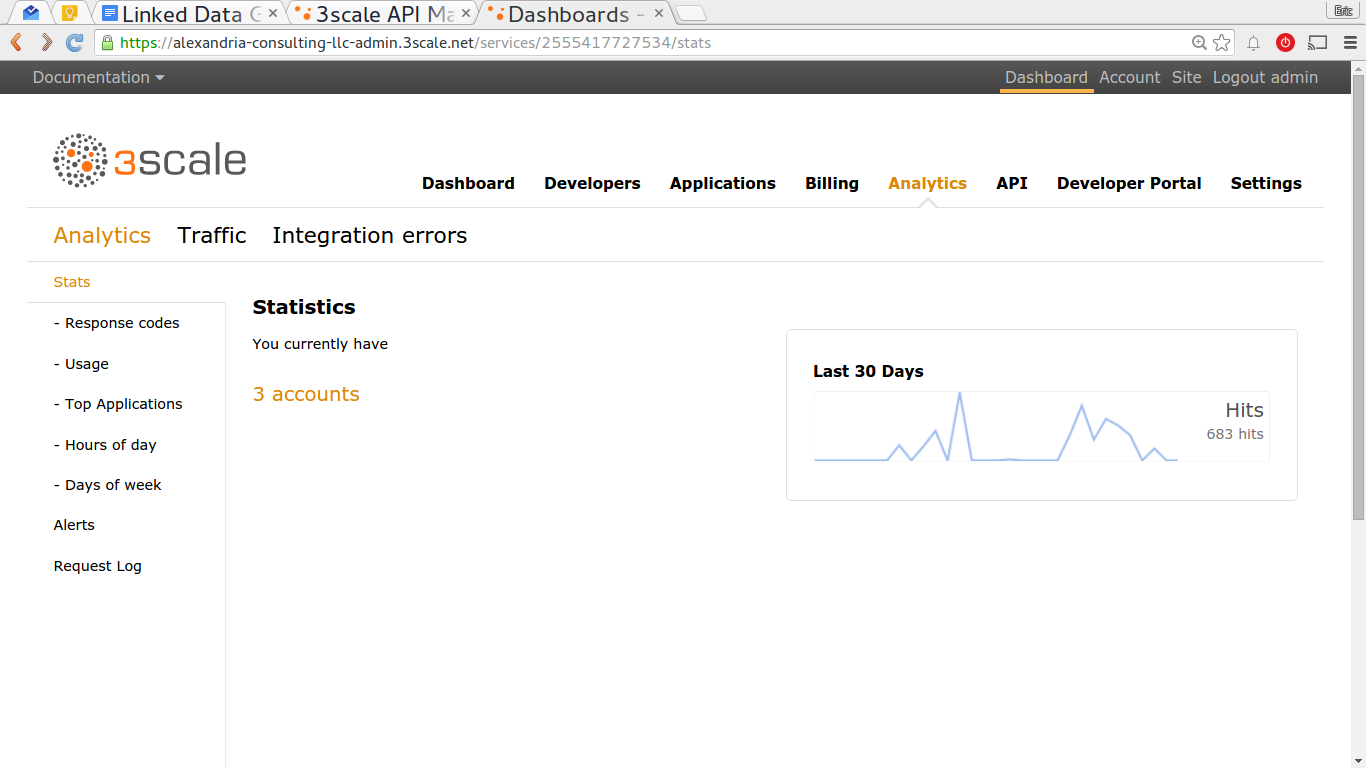
##### Images: Various images of sending deconstructed test messages.

These test messages can come from any test tool and be sent to the controlling web server. If the correct access token is supplied, it will get protected data back, just as the mock web app did.

### Example Analytics On Access To The AIRS Linked Data

Aside from providing security and configurable access plan enforcement, using a Web API Gateway also allows you to see how people are using your data, and what data is most popular at what time of day, etc..

## 



##### Image: See total usage. You can also see usage by app.

##### 

##### hours_of_the_day.pngImage: See what hours of the day your customers are accessing data.

##### 

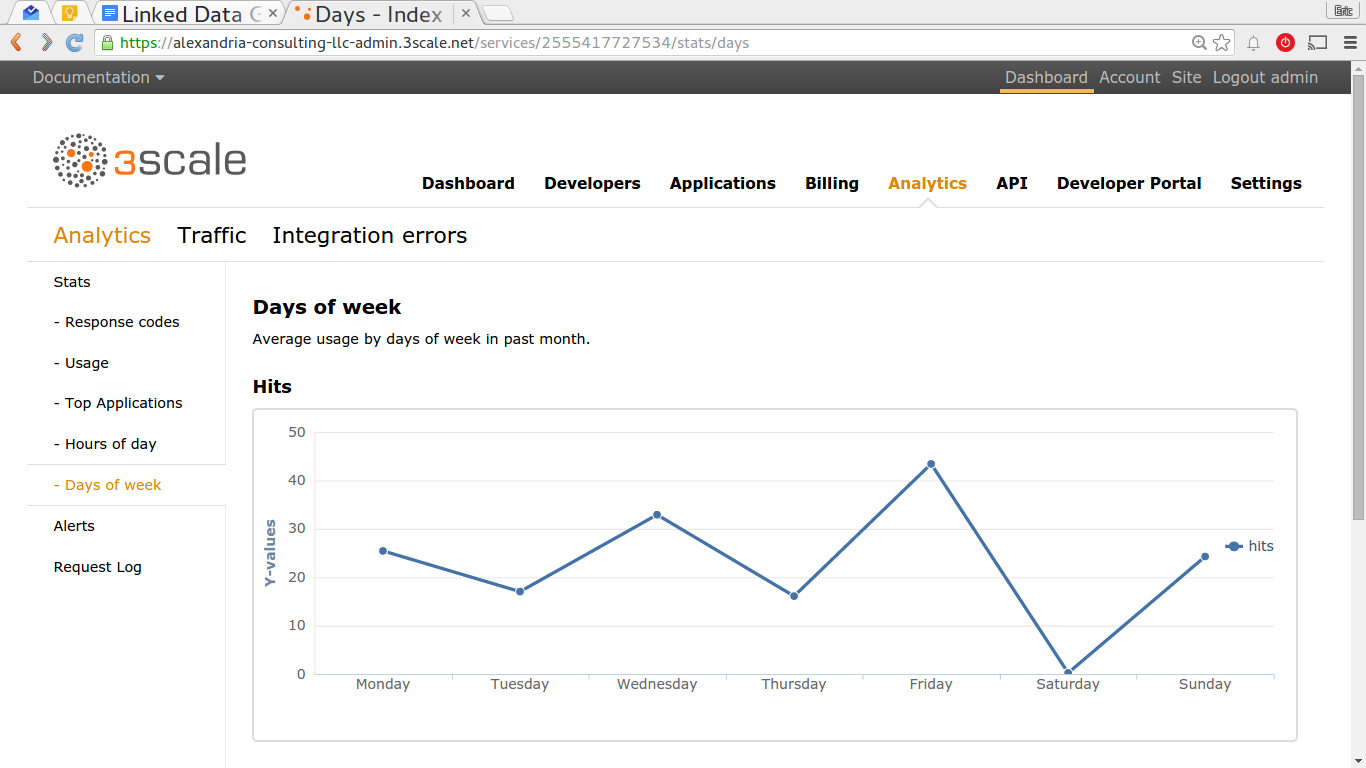
##### 

##### 

￼

##### 

##### 



##### Image: Daily data usage/traffic monitoring.

##### 

##### 

# 

# 

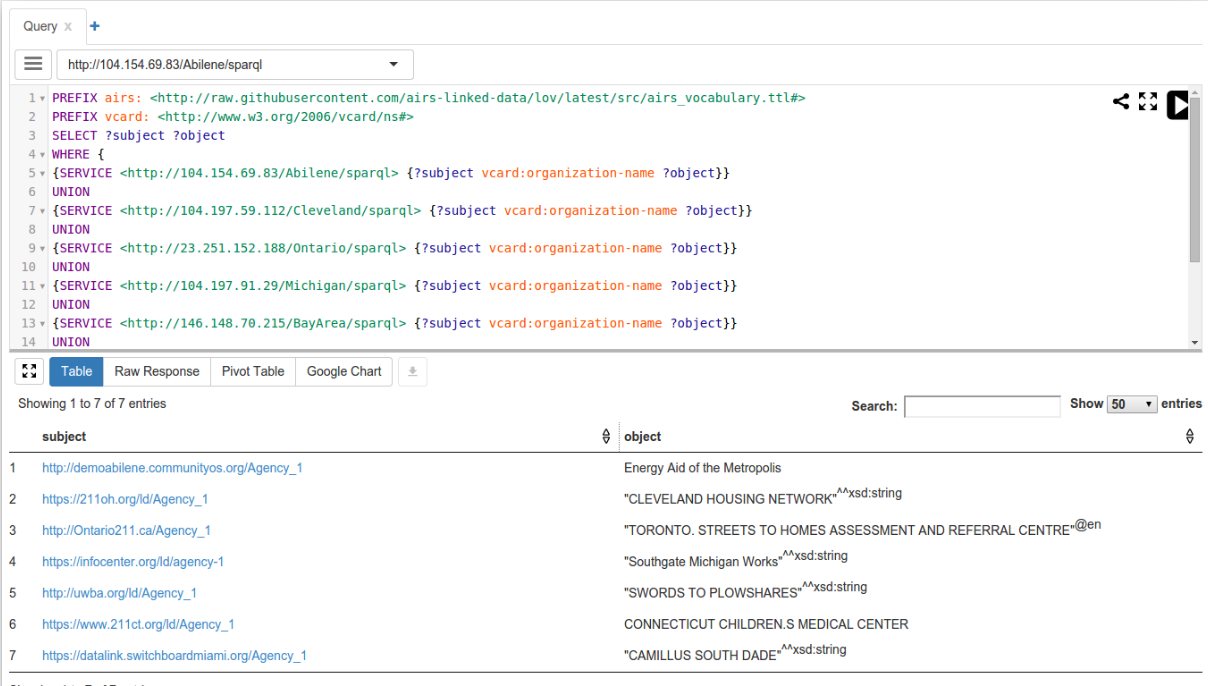
# Appendix 1 - Listing of Pilot Sites

|  |  |  |
| --- | --- | --- |
| **Pilot Site** | **Service Region** | **Home Page Triplestore SPARQL** |
| [211 Abilene](http://www.211texas.org/) | West Central Texas | [home](http://104.154.69.83/), [data](http://104.154.69.83/Abilene/data),[sparql](http://104.154.69.83/Abilene/sparql) |
| [The Information Center](http://theinfocenter.info/) | Southeast Michigan | [url](http://104.197.91.29/), [data](http://104.197.91.29/Michigan/data), [sparql](http://104.197.91.29/Michigan/sparql) |
| [211 United Way of the Bay Area](http://www.uwba.org/about-us) | Bay Area, California | [url](http://146.148.70.215/), [data](http://146.148.70.215/BayArea/data), [sparql](http://146.148.70.215/BayArea/sparql) |
| [211 United Way of Greater Cleveland](http://www.unitedwaycleveland.org/) | Greater Cleveland | [url](http://104.197.59.112/), [data](http://104.197.59.112/Cleveland/data), [sparql](http://104.197.59.112/Cleveland/sparql) |
| [211 United Way of Connecticut](http://www.ctunitedway.org/) | Connecticut | [url](http://104.197.55.108/), [data](http://104.197.55.108/Connecticut/data), [sparql](http://104.197.55.108/Connecticut/sparql) |
| [211 LA County](https://www.211la.org/) | Los Angeles County | (not implemented) |
| [211 Switchboard of Miami](http://switchboardmiami.org/) | Greater Miami | [url](http://104.197.24.64/), [data](http://104.197.24.64/Miami/data), [sparql](http://104.197.24.64/Miami/sparql) |
| [211 Ontario](http://www.211ontario.ca/) | Ontario, Canada | [url](http://23.251.152.188/), [data](http://23.251.152.188/Ontario/data),[sparql](http://23.251.152.188/Ontario/sparql) |

# 

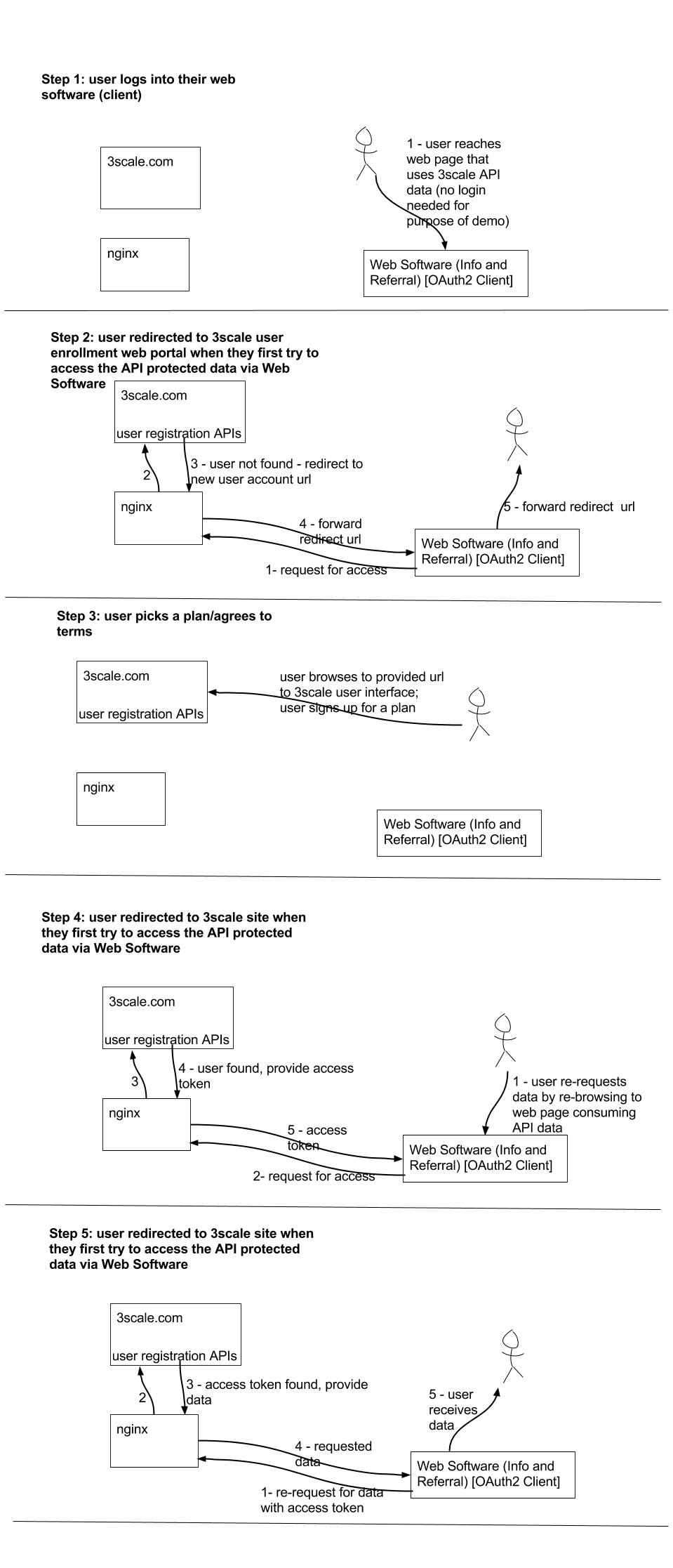
# 

# Appendix 2 - Screenshot Of Example Phase 1 Multi-location Service Lookup



# Appendix 3 - 3rd Party Application Data Access Workflow, using OAuth 2OAuth_Flow.jpg

# 



1. the license is the Creative Commons Zero License for Public Domain works: <https://github.com/airsalliance/lov/blob/master/LICENSE> and <https://creativecommons.org/publicdomain/zero/1.0/> . [↑](#footnote-ref-1)
2. see <http://alexandriaconsulting.com/Members/ejahn/docs/making-information-and-referral-accessible>. [↑](#footnote-ref-2)
3. see <https://www.w3.org/standards/semanticweb/data>. [↑](#footnote-ref-3)
4. see <http://airsnetworker.airs.org/glossary> (may require login credentials). [↑](#footnote-ref-4)
5. see <https://www.w3.org/TR/owl-features/> for a description of the web vocabulary standard used in this demonstration. The AIRS vocabulary using this standard is located at <https://github.com/airsalliance/lov/wiki>. [↑](#footnote-ref-5)
6. see <https://github.com/airsalliance/lov/issues>. [↑](#footnote-ref-6)
7. see <http://airsnetworker.airs.org/communities/community-home?CommunityKey=bac702a3-9be9-4801-9cfa-23f0aafbd332> (may require login). [↑](#footnote-ref-7)
8. see <https://github.com/airsalliance/airs-xml/blob/v3.1.0/trunk/airs.xsd>. [↑](#footnote-ref-8)
9. see <https://github.com/airsalliance/lov/commits/master>. [↑](#footnote-ref-9)
10. Pilots sites and progress are listed here: <https://github.com/airsalliance/linked-data-pilots/wiki/Listing-Of-Pilot-Sites>. [↑](#footnote-ref-10)
11. The extracted data sets are located at <https://docs.google.com/spreadsheets/d/1YWwmgJu5mwlgYLiAPDqf58j1sgsZs9FjX2tNJWonlt8/edit?usp=sharing>. [↑](#footnote-ref-11)
12. Each site’s server location is listed within the “data” links found at <https://github.com/airsalliance/linked-data-pilots/wiki/Listing-Of-Pilot-Sites>, but the servers are not usually left running to control costs. Please contact Alexandria Consulting if you would like any of the servers reactivated. [↑](#footnote-ref-12)
13. see [API](https://en.wikipedia.org/wiki/Application_programming_interface). [↑](#footnote-ref-13)
14. The Gateway service provider matrix we used is located at: <https://docs.google.com/spreadsheets/d/1i0bfIIPk6gSIoVnXO0M9xO8MBn-zFXW6pk_kJ4I2usQ/edit?usp=sharing>. New options such as Amazon.com have emerged since the matrix was developed, but we did have a pleasant experience working with 3scale.net and recommend them as a Gateway service provider. [↑](#footnote-ref-14)
15. Alexandria Consulting has no vested interest in 3scale.net, nor any other Gateway service provider. [↑](#footnote-ref-15)
16. see <http://oauth.net/2/> for more details on OAuth 2. [↑](#footnote-ref-16)
17. The installed scripts are located at <https://github.com/airsalliance/gateway_implementation_demo/tree/master/proxy_configs>. 3scale also provides the following additional workflows to choose from. [↑](#footnote-ref-17)
18. see http://oauthlib.readthedocs.org/en/latest/oauth2/grants/credentials.html for a general discussion of this workflow and see <https://github.com/3scale/nginx-oauth-templates/tree/master/oauth2> for 3scale’s implementation instructions for this workflow. We felt this workflow most fitting for the application scenario of a vendor I&R software accessing outside data on behalf of one of its licensed user accounts. The Client Credentials workflow entails a degree of trust between the I&R vendor application and the I&R application user, which is typically the case. [↑](#footnote-ref-18)
19. see <https://tools.ietf.org/html/rfc6749#section-1.3.4>. [↑](#footnote-ref-19)
20. If the service is not running, please contact Alexandria Consulting to turn it back on. We often mothball services for cost savings, but they can be easily restarted for further demonstration and testing. [↑](#footnote-ref-20)