OLRC

Ontario Library Research Cloud: From Idea to Infrastructure in Three Easy Years

Dale Askey @daskey
Nick Ruest @ruebot

Roadmap

History and background

Technical overview

Current state/Work in progress

Pricing

Governance and participation

Acronyms & names, decoded

– – –

OLRC (oh-lurk) = Ontario Library Research Cloud, built by 11 OCUL partners

OCUL (oh-cool) = Ontario Council of University Libraries - 21 total members

Scholars Portal (SP) = service of OCUL, shared collections and technology infrastructure, hosts/operates OLRC core

History

2012 - idea articulated, original proposal drafted

2013, summer - ready to start, but had a “funding accident”

2014 - purchase, deployment

2015 - final testing, production launch in October

2016 - the fun part starts

Multi-site, widely distributed

~420 km between most distant nodes, as crow flies

Technical overview - Commodity hardware!

– – –

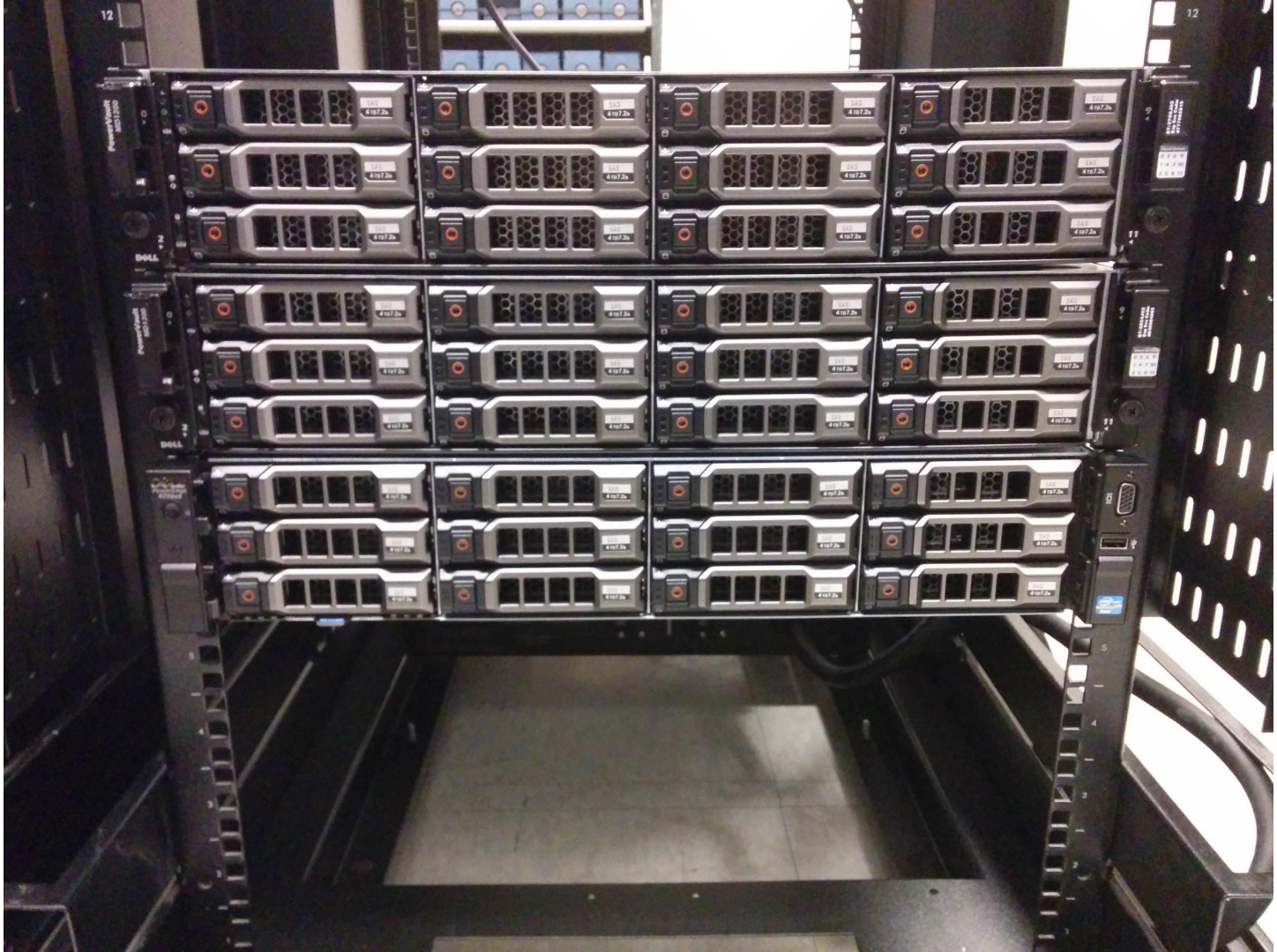
19 Dell PowerEdge R720xd heads

77 MD 1200 storage shelves

Each carries 12x4TB = 48 TB

OpenStack Swift (object storage)

Parent node: SP data centre @ Univ. of Toronto



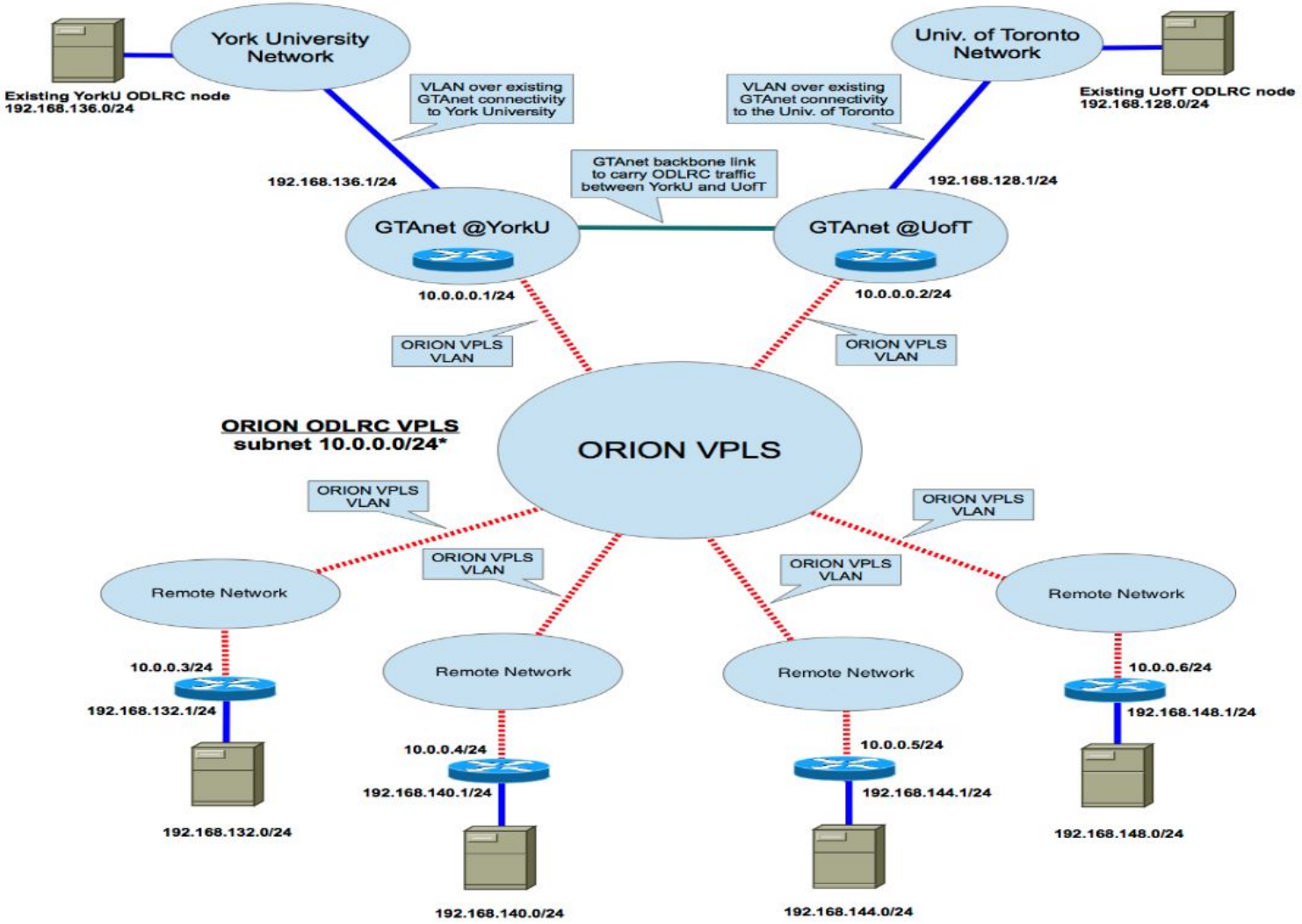
Technical overview

– – –

4 R620 (OpenStack proxy/authentication)

15 10Gbit network switches + fiber optics

ORION (Ontario-based ‘Internet2’) VLAN between nodes


York University Network
Existing YorkU ODLRC node
192.168.136.0/24
VLAN over existing GTAnet connectivity to York University
Univ. of Toronto Network
Existing UofT ODLRC node
192.168.128.0/24
VLAN over existing GTAnet connectivity to the Univ. of Toronto
192.168.136.1/24
GTAnet backbone link to carry ODLRC traffic between YorkU and UofT
192.168.128.1/24
GTAnet @YorkU
GTAnet @UofT
10.0.0.0.1/24
10.0.0.0.2/24
ORION VPLS VLAN
ORION VPLS VLAN
ORION ODLRC VPLS
subnet 10.0.0.0/24*
ORION VPLS
ORION VPLS VLAN
ORION VPLS VLAN
ORION VPLS VLAN
ORION VPLS VLAN
Remote Network
Remote Network
Remote Network
Remote Network
10.0.0.3/24
192.168.132.1/24
192.168.132.0/24
10.0.0.4/24
192.168.140.1/24
192.168.140.0/24
10.0.0.5/24
192.168.144.1/24
192.168.144.0/24
10.0.0.6/24
192.168.148.1/24
192.168.148.0/24

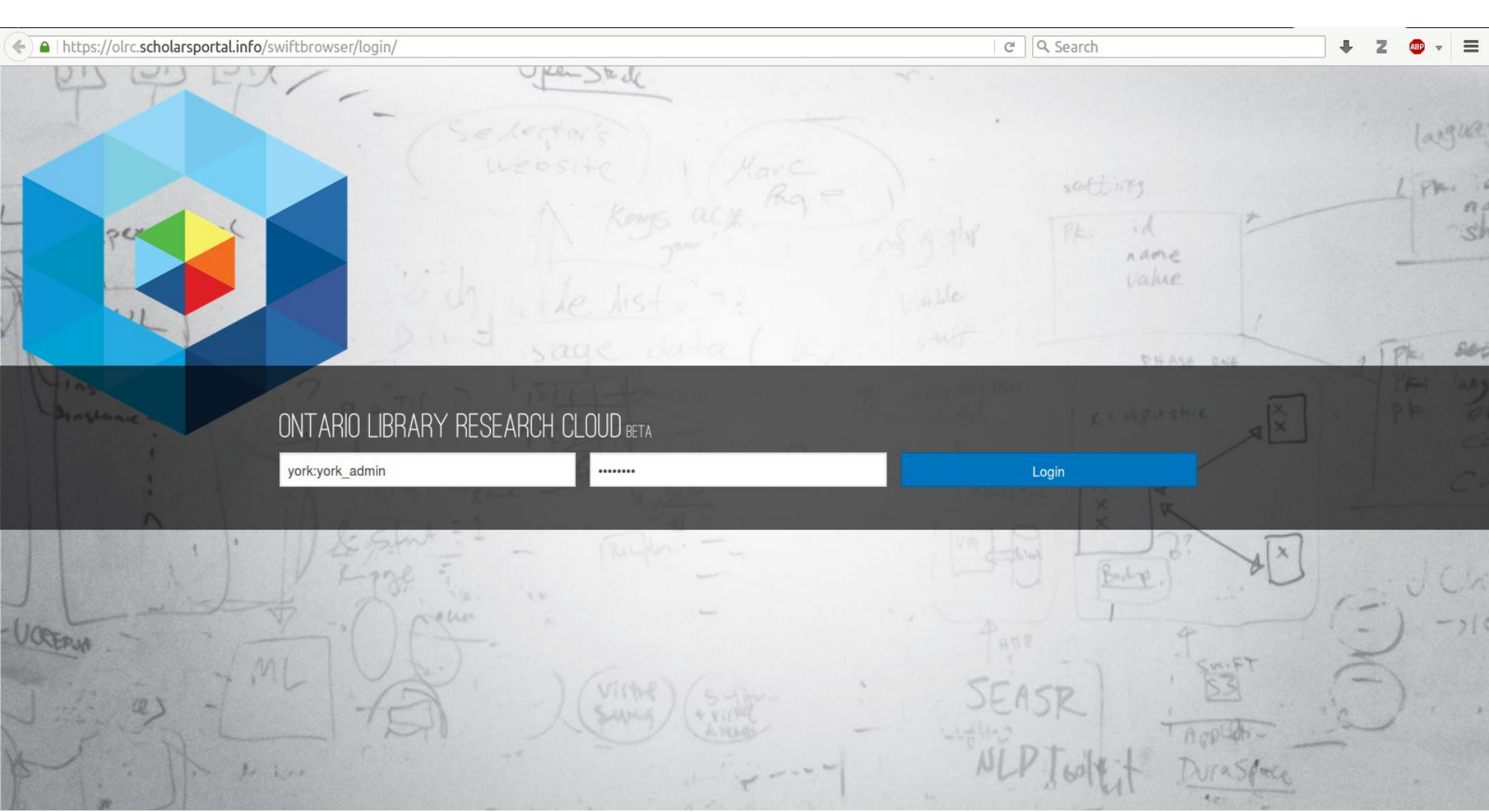
Access

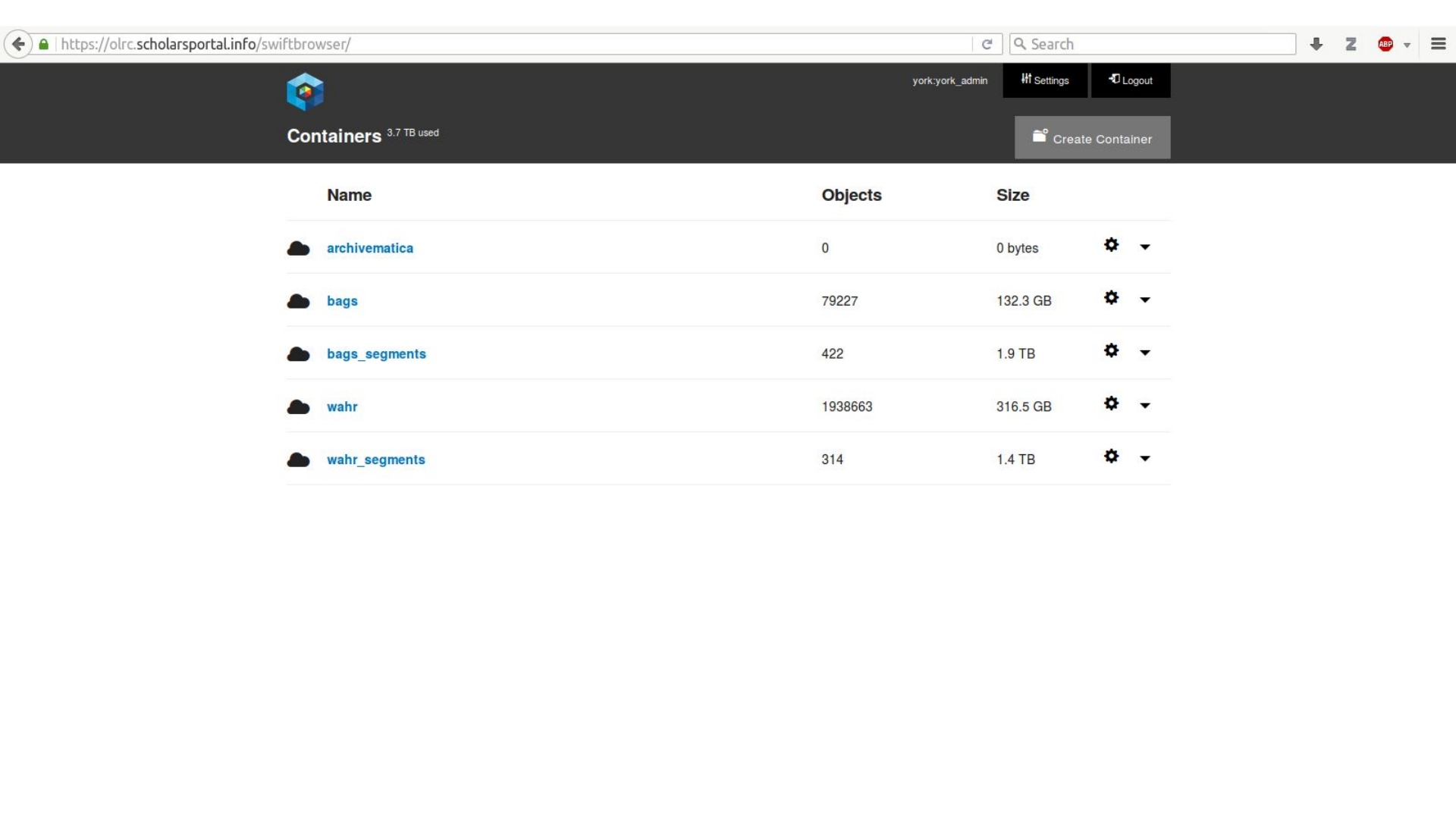
OpenStack Swift API

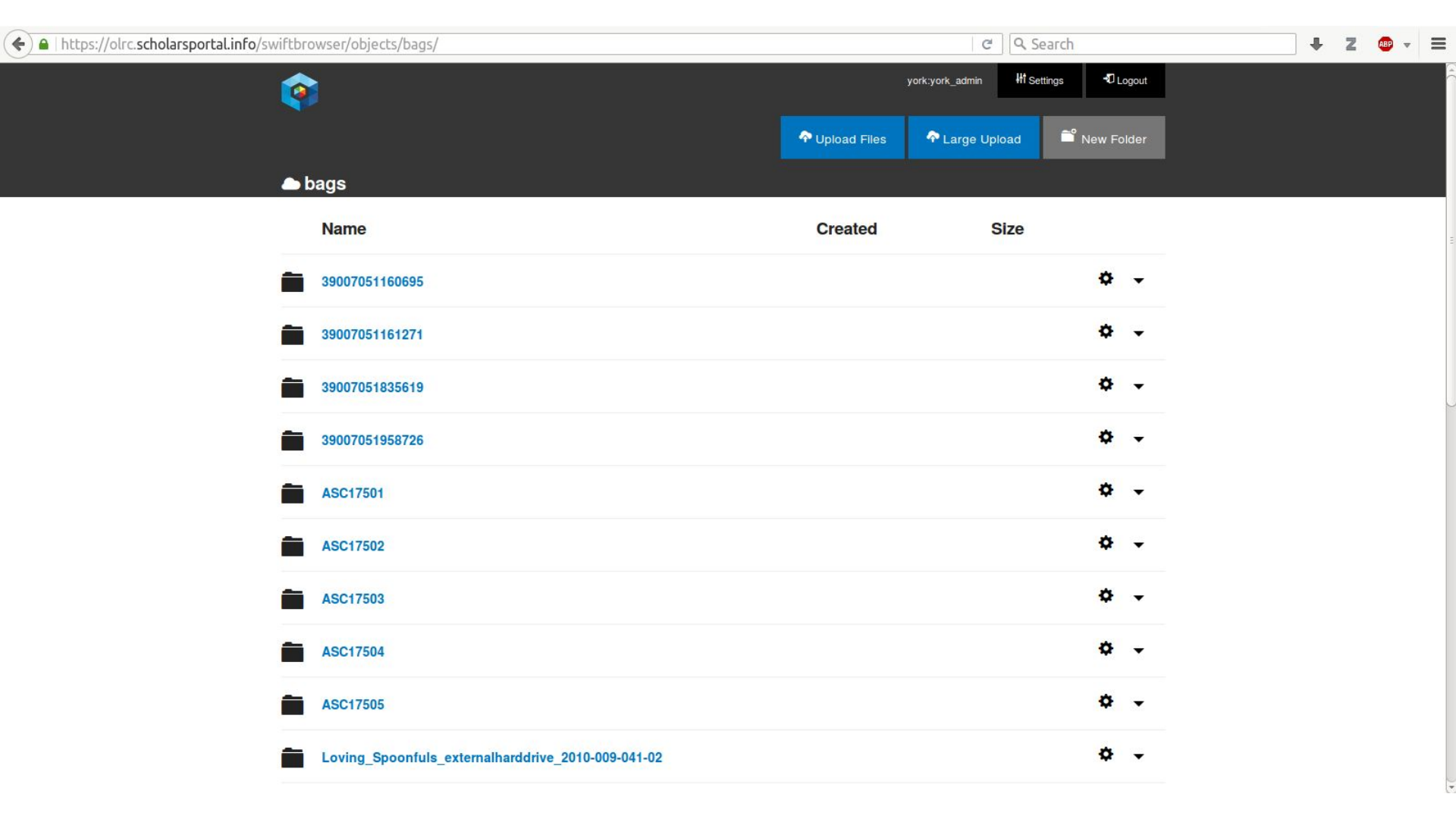
```
[nruest@rho:nruest]$ swift stat
                     Account: AUTH_87b2d1b42b1043388d2649005e7b418c
                  Containers: 5
                     Objects: 2018599
                       Bytes: 4068423484258
Containers in policy "policy-0": 5
   Objects in policy "policy-0": 2018599
     Bytes in policy "policy-0": 4068423484258
             Meta Temp-Url-Key: YjAyMTJm
    X-Account-Project-Domain-Id: default
                  X-Timestamp: 1444062593.16277
                   X-Trans-Id: tx299c84bbd95444a2a6013-0056d6d67e
                Content-Type: text/plain; charset=utf-8
               Accept-Ranges: bytes
[nruest@rho:nruest]$ swift list
archivematica
bags
bags_segments
wahr
wahr_segments
[nruest@rho:nruest]$ swift list bags
39007051160695/bag-info.txt
39007051160695/bagit.txt
39007051160695/data/39007051160695-MARC.xml
39007051160695/data/39007051160695-TECHNICAL.xml
39007051160695/data/39007051160695.avi
39007051160695/data/39007051160695.pdf
39007051160695/manifest-md5.txt
39007051161271/bag-info.txt
39007051161271/bagit.txt
39007051161271/data/39007051161271-MARC.xml
39007051161271/data/39007051161271-TECHNICAL.xml
39007051161271/data/39007051161271.avi
39007051161271/data/39007051161271.pdf
39007051161271/data/Thumbs.db
39007051161271/manifest-md5.txt
39007051835619/bag-info.txt
39007051835619/bagit.txt
39007051835619/data/39007051835619-TECHNICAL.xml
39007051835619/data/39007051835619.avi
39007051835619/data/39007051835619.pdf
39007051835619/data/Thumbs.db
39007051835619/manifest-md5.txt
39007051958726/bag-info.txt
39007051958726/bagit.txt
39007051958726/data/39007051958726-MARC.xml
39007051958726/data/39007051958726-TECHNICAL.xml
39007051958726/data/39007051958726.avi
```

```
[nruest@rho:data_sets]$ rclone copy elxn42 olrc:wahr/elxn42
2016/03/02 07:05:38 Swift container wahr path elxn42/: Building file list
2016/03/02 07:05:38 Swift container wahr path elxn42/: Waiting for checks to finish
2016/03/02 07:06:37
Transferred:   3136947015 Bytes (50673.85 kByte/s)
Errors:                 0
Checks:                 0
Transferred:           24
Elapsed time:      1m0.4s
Transferring:
 *                          LAC/elxn42-LAC.json:  4% done. avg: 14551.2, cur: 13964.6 kByte/s. ETA: 21m30s
 *           byDate/elxn42-tweets-20150806.json: 100% done. avg: 13795.2, cur: 13809.9 kByte/s. ETA: 0s
 *           byDate/elxn42-tweets-20150817.json: 39% done. avg: 11282.0, cur: 11072.7 kByte/s. ETA: 8s
 *           byDate/elxn42-tweets-20150818.json: 25% done. avg: 11028.5, cur: 10192.7 kByte/s. ETA: 15s

2016/03/02 07:07:37
Transferred:   6125276999 Bytes (49659.85 kByte/s)
Errors:                 0
Checks:                 0
Transferred:           60
Elapsed time:      2m0.4s
Transferring:
 *                          LAC/elxn42-LAC.json:  8% done. avg: 13979.3, cur: 12150.6 kByte/s. ETA: 23m37s
 *           byDate/elxn42-tweets-20150922.json: 53% done. avg: 9653.0, cur: 9618.7 kByte/s. ETA: 6s
 *           byDate/elxn42-tweets-20150923.json: 42% done. avg: 9650.7, cur: 10009.2 kByte/s. ETA: 9s
 *           byDate/elxn42-tweets-20150924.json: 15% done. avg: 9052.3, cur: 8865.4 kByte/s. ETA: 30s

2016/03/02 07:08:37
Transferred:   9195629629 Bytes (49764.04 kByte/s)
Errors:                 0
Checks:                 0
Transferred:           69
Elapsed time:      3m0.4s
Transferring:
 *                          LAC/elxn42-LAC.json: 13% done. avg: 13881.3, cur: 14011.6 kByte/s. ETA: 19m30s
 *           byDate/elxn42-tweets-20151001.json: 97% done. avg: 14376.8, cur: 13786.0 kByte/s. ETA: 0s
 *           byDate/elxn42-tweets-20151002.json: 30% done. avg: 13920.6, cur: 15009.4 kByte/s. ETA: 20s
 *           byDate/elxn42-tweets-20151003.json: 34% done. avg: 13043.9, cur: 12854.3 kByte/s. ETA: 15s

2016/03/02 07:09:37
Transferred:   12237965024 Bytes (49702.44 kByte/s)
Errors:                 0
Checks:                 0
Transferred:           76
Elapsed time:      4m0.4s
Transferring:
 *                          LAC/elxn42-LAC.json: 17% done. avg: 13496.9, cur: 11183.9 kByte/s. ETA: 23m20s
 *           byDate/elxn42-tweets-20151008.json: 50% done. avg: 12280.7, cur: 11342.1 kByte/s. ETA: 16s
 *           byDate/elxn42-tweets-20151009.json: 18% done. avg: 11237.5, cur: 14308.7 kByte/s. ETA: 23s
 *           byDate/elxn42-tweets-20151010.json: 15% done. avg: 10295.9, cur: 10329.9 kByte/s. ETA: 30s
```

Swift Browser


https://olrc.scholarsportal.info/swiftbrowser/login/
Search
ONTARIO LIBRARY RESEARCH CLOUD BETA
york:york_admin
Login


https://olrc.scholarsportal.info/swiftbrowser/
Search
york:york_admin
Settings
Logout
Containers 3.7 TB used
Create Container
Name
Objects
Size
archivematica
0
0 bytes
bags
79227
132.3 GB
bags_segments
422
1.9 TB
wahr
1938663
316.5 GB
wahr_segments
314
1.4 TB


https://olrc.scholarsportal.info/swiftbrowser/objects/bags/
Search
york:york_admin
Settings
Logout
Upload Files
Large Upload
New Folder
bags
Name
Created
Size
39007051160695
39007051161271
39007051835619
39007051958726
ASC17501
ASC17502
ASC17503
ASC17504
ASC17505
Loving_Spoonfuls_externalharddrive_2010-009-041-02

Current state

– – –

1.2 PB (net)

300 TB dedicated for original 11 partners plus SP

20 TB container for each partner institution

-- ***utilization is poor, thus far*** --

~1 PB - U of Toronto to support digital scholarship & other projects - *not geographically replicated*

Work in progress - York

Bag storage (Dark archive)

Archivematica integration

Fedora 4 integration

Bacula

Partnerships with Central IT w/research life cycle

Work in progress - McMaster

McMaster: 14 TB, but in testing phase

> 7TB raw archives, 6TB Digital Archive, 1 TB IR & OJS

Bit of trial and error with tools, learning to use object storage

Not preservation, but working toward it through testing and development

Costs: some good news

Scholars Portal: established and capable

Dedicated partners; others ready to step in

OCUL governance model and history

Costs: who and how much?

Who:

- original partners
- other OCUL institutions
- other non-Ontario institutions
- researchers

Costs: model needed

– – –

Relied extensively on Ratliff and Goldstein (2010):

> *DataSpace: A Funding and Operational Model for Long-Term Preservation and Sharing of Research Data*

http://arks.princeton.edu/ark:/88435/dsp01w6634361k

Costs: assume it's going to rain

Calculated in partner costs, even if those are not currently levied (hardware refresh, data centre costs, bandwidth, etc.)

Conservative estimate of annual decrease in storage costs

Five-year hardware refresh cycle

1 CAD != 1 €

1 CAD = .79 USD, .69 EUR, .54 GBP

Estimated per TB costs, OCUL members

Annual subscription: $320/TB*

3% discount for 10TB block

6% discount for 50TB block

* OCUL schools pay a $1600/yr 5TB minimum

Estimated per TB costs, non-OCUL members

Annual subscription: $790/TB

3% discount for 10TB block

6% discount for 50TB block

* Non-OCUL clients pay a $3950/yr 5TB minimum

Costing questions in search of answers

When do we let others access the storage?

Who, i.e.- how broadly do we throw open the door?

Do we sell directly, or via partners?

Will others replicate our model?

Costing: the big question

What should a TB of preservation-capable storage cost? $10? $4000?

Governance

Project Phase

* Steering Committee

* Technical Committee

Governance

Production Phase

* OCUL Cloud Storage Committee
    * One member from each institution

OCUL Cloud Storage Committee

Supports the operation and future development of the the Storage Cloud.

Makes recommendations to the OCUL Executive regarding the plans, mission, goals, policies, operations, and development of the Storage Cloud, including possible partnerships with non--OCUL institutions.

OCUL Cloud Storage Committee

Facilitates informed decision -making on the part of current and future members.

Develops and reviews annually a storage pricing model for OCUL members and non--OCUL clients.

Considers, reviews, and approves usage and operational policies for the Storage Cloud as recommended by Scholars Portal staff.

Governance/Collaboration - issues

Tools - we need to develop tools and documentation for existing tools

Repository / Archivematica integration

Governance/Collaboration - issues

Contributing to work, not just riding along

Developing expertise, diving in

Overemphasis on policy

Concerned about scenarios that may not materialize

Last thoughts

We collaborated to build it; now we need to collaborate on using it and developing tools

Need to let others use it sooner rather than later

Need to solidify future governance within OCUL

Thank you

Much credit due to:

- Alan Darnell, Amaz Taufique, Steve Marks
- Partner libraries
- OLRC Admin & Tech Committees
- SP/U of Toronto systems teams: Graham Stewart, Steve Baroti, Chris Crebolder, Miki Wong, Harpinder Singh, Bikram Singh