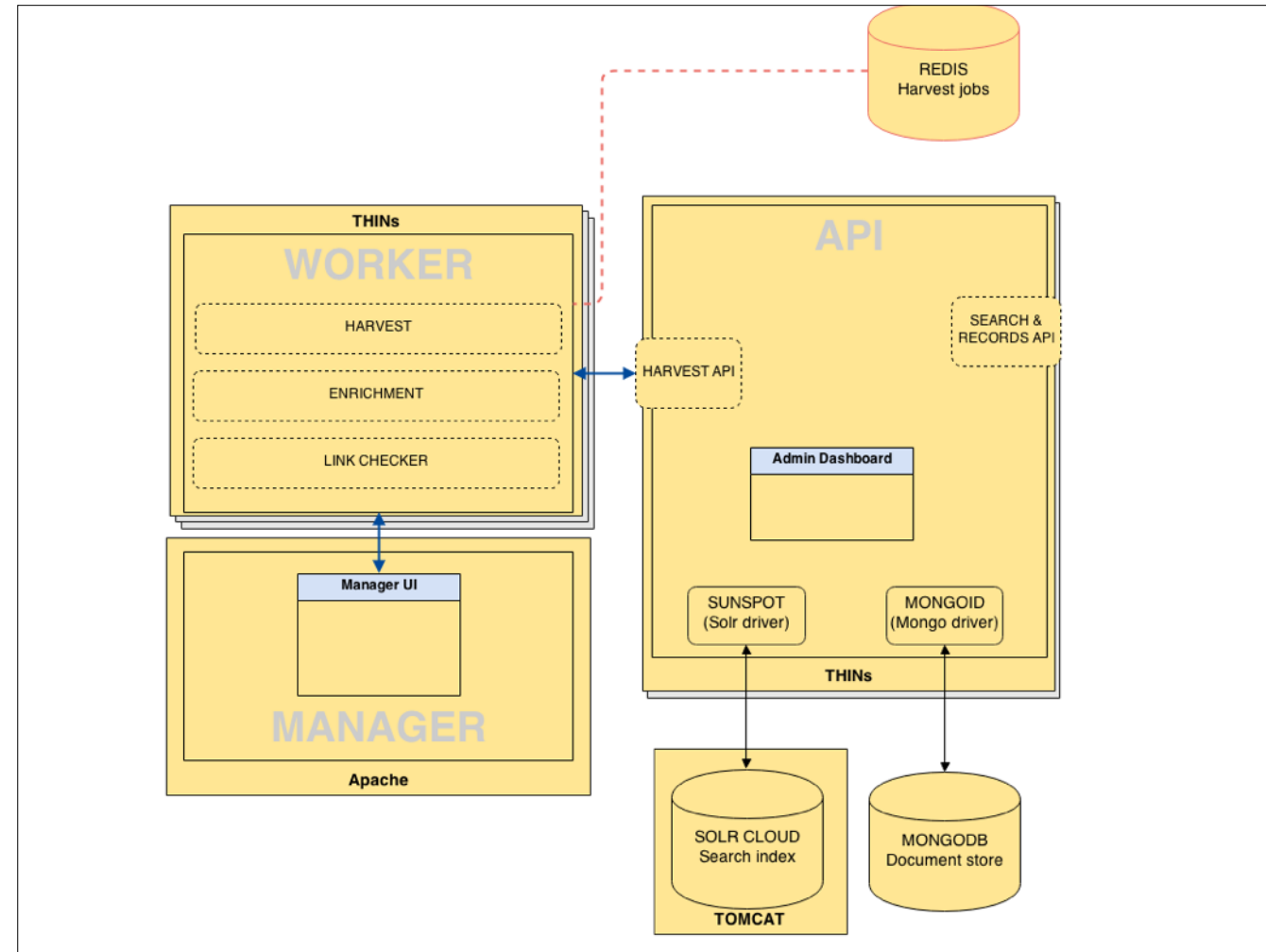




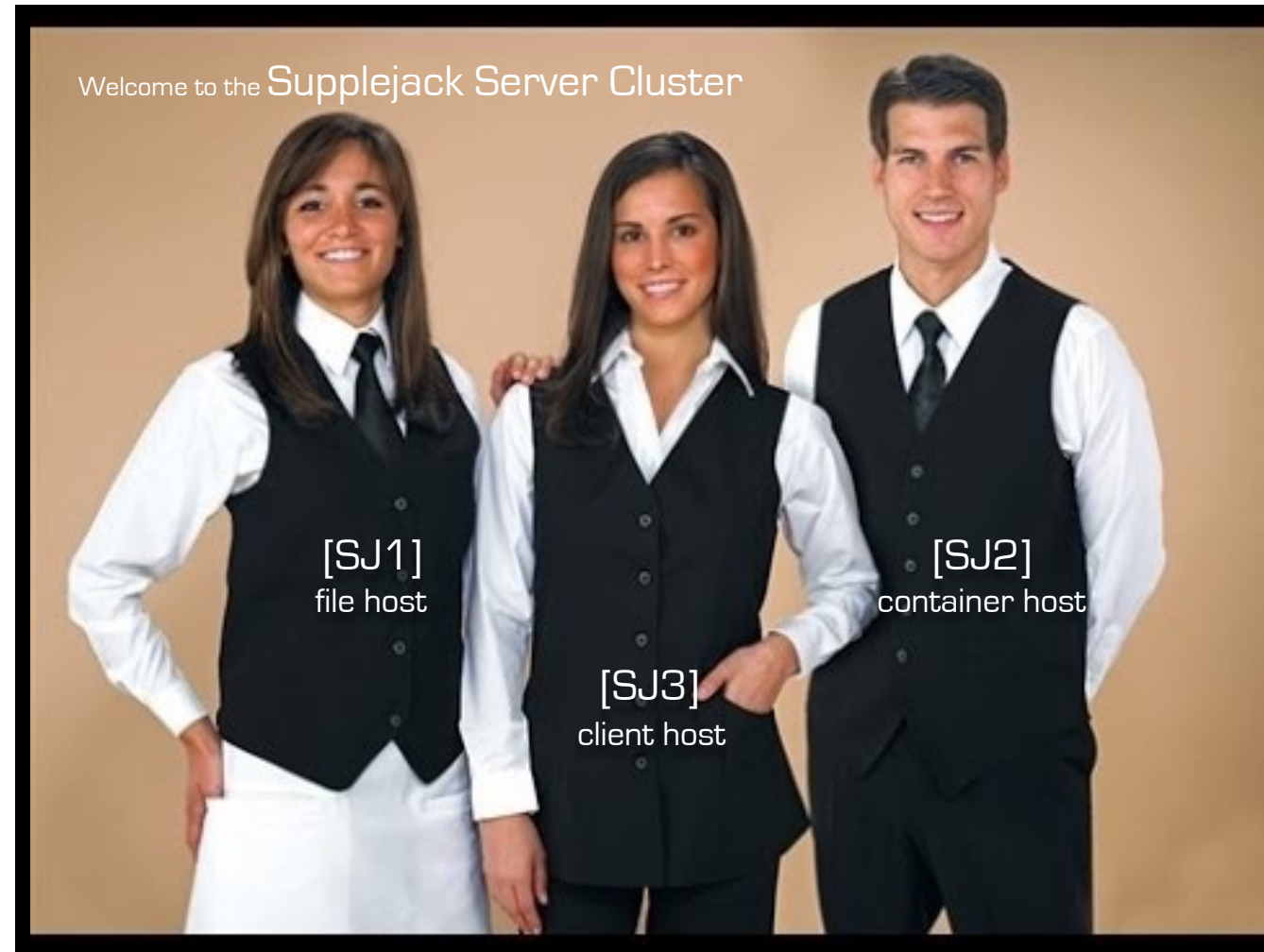
Matt and I been working together (remotely) on this project for about four months (since December), this is an brief overview some of that work  
you can contact us via increasingly secure channels



We have been working primarily with SuppleJack - the technology developed by Digital NZ, used to power their national digital library

this is a brief overview of Supplejack system architecture

For folks that aren't interested in the back end (and even if you are) the Important part here is that is has a user interface where you do the bulk of your work: harvesting, ingesting, mapping, scheduling . . .



Overview of the system architecture we have implemented for our instance.

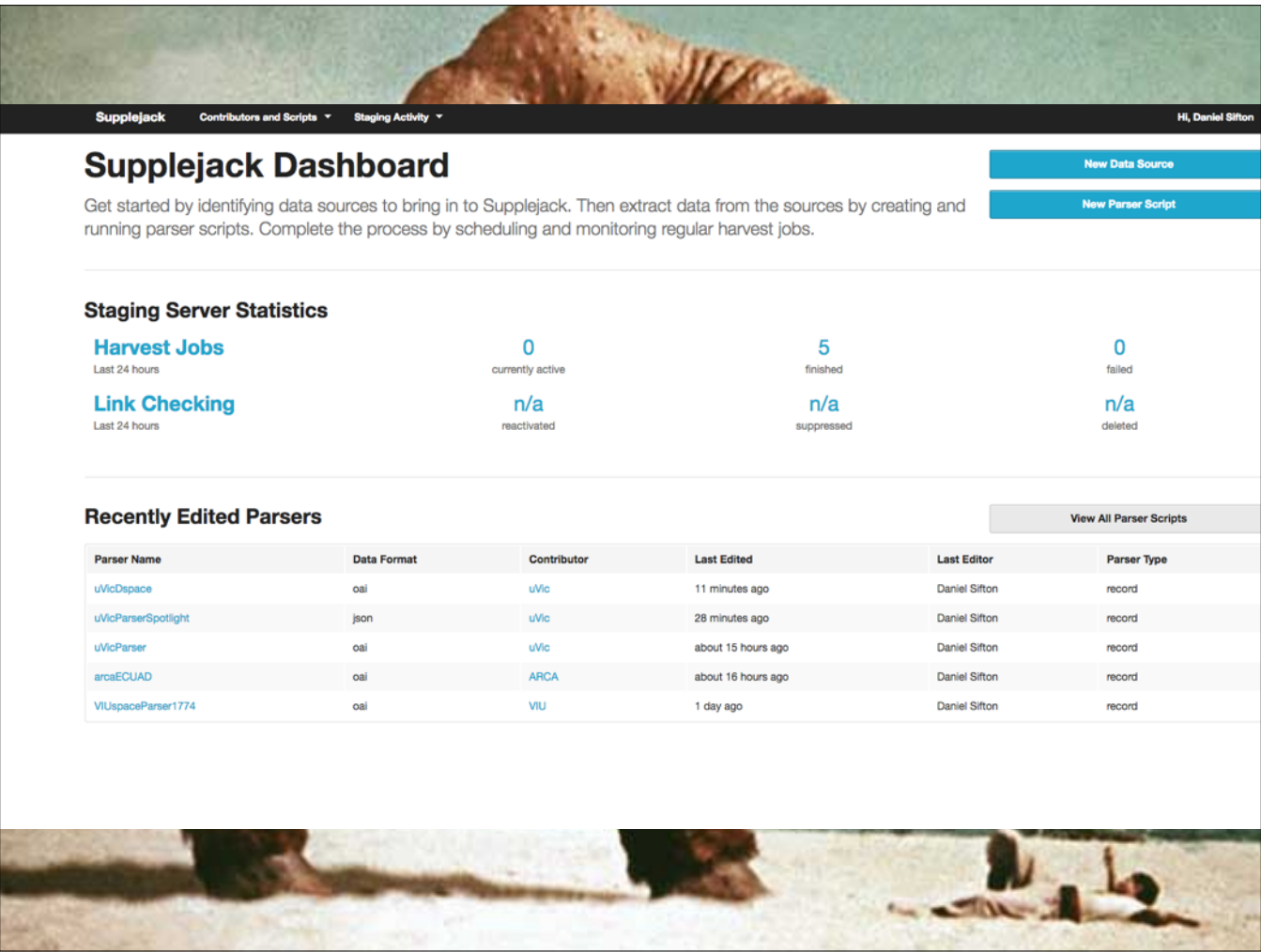
We're working with docker here and these friendly faces are a visualization of how we structure our docker files/containers and our discovery layer in the cloud

SJ1 [the file host]- we host the files for the docker application

SJ2 [the container host] - we host the docker containers (think of this as 7 lightweight VMs)

SJ3 [the client host] - SJ2 puts out a machine readable API that is then translated by the front-end [SJ3] (the client) for human beings.

welcome to the supplejack server cluster!



The hotseat, is the Supplejack Manager (application) dashboard where we get an overview of recent activities and via the top menu: , schedule harvest jobs, create users, define link checking rules, view errors . . . lots of admin things the most important of which is building parser scripts (that we use to map metadata from ingested records)

# Parser Scripts

New Parser Script

Search:

Harvest Name	Data Format	Partner	Source ID	Last Updated	Last Edited By	Parser Type
uVicDspace	oai	uVic	uVic	2017-04-10 20:27:19 +0000	Daniel Sifton	record
uVicParserSpotlight	json	uVic	uVic	2017-04-10 20:10:52 +0000	Daniel Sifton	record
uVicParser	oai	uVic	uVic	2017-04-10 05:43:56 +0000	Daniel Sifton	record
arcaECLUAD	oai	ARCA	ARCA	2017-04-10 04:33:40 +0000	Daniel Sifton	record
VIUspaceParser1774	oai	VIU	VIUspace5	2017-04-09 19:43:05 +0000	Daniel Sifton	record
VIUspaceBase	oai	VIU	VIUspace5	2017-04-08 07:31:33 +0000	Daniel Sifton	record
VIUspaceParser2696	oai	VIU	VIUspace5	2017-04-08 06:32:18 +0000	Daniel Sifton	record
Waterloo OAI	oai	Waterloo	Waterloo2	2017-04-08 04:34:22 +0000	Daniel Sifton	record
VIU440	oai	TestContributor	TestSource	2017-04-07 23:42:46 +0000	Daniel Sifton	record
VIUspaceGeoParser	oai	VIU	VIUspace3	2017-04-07 19:09:18 +0000	Daniel Sifton	record
DigitalNZ	json	NZ	NZdata	2017-04-06 21:45:49 +0000	Daniel Sifton	record
UBCparser	json	UBC	UBC	2017-04-06 11:21:48 +0000	Matthew Barry	record
SFUparser	oai	SFU	SFU	2017-04-05 16:18:39 +0000	Daniel Sifton	record
SqPLparser	xml	Squamish	SqPL	2017-04-05 05:41:53 +0000	Daniel Sifton	record
ArcheionParser	oai	AtoMArchives	AtoMarchives	2017-04-04 20:40:27 +0000	Daniel Sifton	record
VIUthumbs	xml	VIU	VIUspace3	2017-04-04 06:12:49 +0000	Daniel Sifton	record
VIUyoutube	json	VIU2	VIUspace4	2017-04-03 21:08:24 +0000	Daniel Sifton	record
arcaTWU	oai	ARCA	ARCA	2017-04-03 20:47:48 +0000	Daniel Sifton	record
arcaUNBC	oai	ARCA	ARCA	2017-04-03 20:18:24 +0000	Daniel Sifton	record
WPLparser2	xml	WPL	WPL2	2017-04-03 17:41:10 +0000	Daniel Sifton	record

Showing 1 to 20 of 35 entries

more parser scripts

SupplejackContributors and ScriptsStaging Activity

Hi, Daniel Sifton

WPLparser2

record

```
1 class WPLparser2 < SupplejackCommon::Xml::Base
2
3
4   base_url "http://159.203.21.122/data/WestVanPub.xml"
5
6
7   record_selector "//doc"
8   record_format :xml
9
10  attributes :display_content_partner, default: "West Vancouver Memorial Library"
11
12
13  attributes :title, xpath: "//field[@name='title']"
14
15    reject if do
16      get(:title).find_with(/Collection/).present?
17    end
18
19
20  attributes :subject, xpath: "//field[@name='subject']"
21
22  attributes :category, xpath: "//field[@name='type']"
23
24  attributes :publisher, xpath: "//field[@name='publisher']"
25
26  attributes :display_date, xpath: "//field[@name='dateOldest']"
27
28
29
30  attribute :creator, xpath: "//field[@name='creator']" # do
31    #tag = get(:creator).split(",")
32    #compose(tag.select(:last), " ", tag.select(:first))
33  #end
34
35
36  attributes :subject, xpath: "//field[@name='subject']"
37
38  attributes :description, xpath: "//field[@name='description']", truncate: 170 do
39    get(:description).first
40  end
41
42
43
44
45
46  attribute :locAll do
47    fetch("//field[@name='itemLatLong']").mapping( /"/ => "north=", /~/ => " west=")
48  end
49
50
51
52  attribute :locations do
53    locations = get(:locAll).to_a
54    locations.collect do |location|
55      if location.include? "north="
56        lat = /(?!<north=).*?(?=(\,|\s))/match(location)[0].to_f
57      end
58      if location.include? "west="
59        lng = /(?!<west=).*?(?=(\,|\s))/match(location)[0].to_f * -1
60      end
61      # if location.include? "name="
62      #   name = /(?!<name=).*?(?=(\,|\s))/match(location)[0]
63      # end
64      {lat:lat or lng:nil} ? next : {lat: lat, lng: lng}
65    end
66  end
67
68
69
70  attributes :language, xpath: "//field[@name='language']", mappings: {
71    "eng" => "en"
72  }
73
74
75
76  attribute :rights, xpath: "//field[@name='rightsCreativeCommons']", mappings: {
77    "by" => "CC-BY"
78  }
79
80
```

Preview

Current version: StandardLibNames

Edit current version

History

StandardLibNames

04 Apr 05:41 by Daniel

S

WPL2baseCreator

29 Mar 05:57 by Daniel

WPL2base

28 Mar 11:03 by Daniel

Applied parser template.

28 Mar 10:41 by Daniel

Rename Parser

Change Data Source

Delete Parser Script

Disable Full & Flush harvest mode

Here's a whirlwind view of how we write a parser -

define you target (a file - if you can't harvest or define a url)

name the attribute (in green) ID its source field in the harvest 'target'

these attributes are defined in a highly configurable schema - that I'm not showing you

we wrangle the data as we ingest - truncation, substitution, composition, conditionals, exclusions . . .

I'm greatly simplifying this process to allow for time constraints and this is a relatively clean parser. If the data is more complex, it may require a more complex parser,

This is an XML harvest for West Van

SupplejackContributors and ScriptsStaging Activity

Hi, Daniel

UBCICYoutube

1class UBCICYoutube < SupplejackCommon::Json::Base23456789101112131415161718192021222324252627282930313233343536373839404142434445464748495051525354555657585960616263end

Message

Update Parser Script

Preview

Current version: UBCICSubjectmodsMa

Edit current version

History

UBCICSubjectmodsMaps328 Mar 09:14 by Daniel

UBCICSubjectmodsMaps228 Mar 09:13 by Daniel

UBCICSubjectmodsMaps28 Mar 09:05 by Daniel

UBCICDateMod26 Mar 18:35 by Daniel

UBCICCleanUpSubjects23 Mar 19:00 by Daniel

UBCICCleanUpDate23 Mar 06:37 by Daniel

Renamed parser class07 Mar 08:49 by Daniel

UBCICBaseParser07 Mar 08:43 by Daniel

Applied parser template.07 Mar 08:39 by Daniel

Rename Parser

Change Data Source

Delete Parser Script

Disable Full & Flush harvest mode

and here's a parser that harvests JSON from the UBCIC youtube channel - you can see our versioning on the right

WPLparser2

record

Preview

Current version: StandardLibNames

1class WPLparser2 < SupplejackCommon::Xml::Base

2

3

4

5

6be

7

8re

9re

10at

11

12

13at

14

15

16

17

18end

19

20

21at

22

23

24at

25

26at

27

28

29

30at

31

32

33

34

35

36at

37

38at

39

40end

41

42

43

44

45

46at

47

48end

49

50

51

52at

53lo

54l

55

56

57

58

59

60

61

62

63

64

65on

66

67

68

69

70at

71

72

73

74

75

76at

77

78

79

80

81at

82

Previewing records

Status: Preview complete.

Source Data

Harvested Attributes

API Record

< previous - Preview Record - next >

1<doc>

2

3<field name="id">OOL.2822428</field>

4

5<field name="localid">2822428</field>

6

7<field name="accessionNo">AV024</field>

8

9<field name="url">http://digital.westvanlibrary.ca/2822428/data</field>

10

11<field name="title">[Opening ceremonies, West Vancouver Memorial Library, CKWX, November 11, 1950]</field>

12

13<field name="titleSort">opening ceremonies west vancouver memorial library ckwx november 11 1950 2822428</field>

14

15<field name="publisher">CKWX</field>

16

17<field name="description">Audio recording of a delayed radio broadcast of the opening ceremony and dedication of the war memorial library in West Vancouver, on

18

19<field name="description">Title from label on original reel-to-reel recording.</field>

20

21<field name="fulltext">Introduction by radio announcer in studio (00:01); introduction by radio announcer Lori (?) Irving, on-site at the opening ceremony (00:1

22

23<field name="fulltext">Audio (MP3) </field>

24

25<field name="multiPart">>false</field>

26

27<field name="dateSort">19501111</field>

28

29<field name="fDateCentury">1900</field>

30

31<field name="fDateDecade">1950</field>

32

33<field name="fDateYear">1950</field>

34

35<field name="fDateMonth">11</field>

36

37<field name="fDateDay">11</field>

38

39<field name="dateOldest">1950</field>

40

41<field name="dateNewest">1950</field>

42

43<field name="created">2014-08-12T14:17:26Z</field>

44

45<field name="modified">2015-02-24T14:41:15Z</field>

46

47<field name="madePublic">2014-11-04T13:29:29Z</field>

48

49<field name="itemType">Sound recordings</field>

50

51<field name="type">audio</field>

52

as we write a parser, we can preview results:  
from the source file



WPLparser2

record

Preview

Current version: StandardLibNames

1class WPLparser2 < SupplejackCommon::Xsl::Base

2

3

4

5ba

6

7

8re

9re

10

11at

12

13

14at

15

16

17

18end

19

20

21at

22

23

24at

25

26

27at

28

29

30at

31

32

33

34

35

36at

37

38at

39

40end

41

42

43

44

45

46at

47

48end

49

50

51

52at

53loc

54l

55

56

57

58

59

60

61

62

63

64

65

66en

67

68

69

70at

71

72

73

74

75

76at

77

78

79

80

81at

82

Previewing records

Status: Preview complete.

Source DataHarvested AttributesAPI Record

< previous - Preview Record - next >

1{

2"priority": 0,

3"match\_concepts": null,

4"display\_content\_partner": {

5"West Vancouver Memorial Library"

6

7},

8"title": {

9"[Opening ceremonies, West Vancouver Memorial Library, CKWX, November 11, 1950]"

10

11},

12"subject": {

13"Building dedications",

14"Monuments & memorials"

15

16},

17"category": {

18"audio"

19

20},

21"publisher": {

22"CKWX"

23

24},

25"display\_date": {

26"1950"

27

28},

29"creator": {

30

31},

32"description": {

33"Audio recording of a delayed radio broadcast of the opening ceremony and dedication of the war memorial library in West Vancouver, on November 11, 1950, not"

34

35},

36"locAll": {

37

38},

39"locations": {

40

41},

42"language": {

43"en"

44

45},

46"rights": {

47

48},

49"thumbnail\_url": {

50"http://images.ourontario.ca/Partners/WVNL/WVNL002822428t.jpg"

51

52},

53"source\_url": {

54"http://digital.westvanlibrary.ca/2822428/data"

55

56},

57"landing\_url": {

58"http://digital.westvanlibrary.ca/2822428/data"

59

60},

61"internal\_identifier": {

62"http://digital.westvanlibrary.ca/2822428/data"

63

64},

65}

from our harvested attributes



Run Staging Harvest

StandardLibNames

Mode

☒ Normal

☐ Full And Flush

How many records do you wish to harvest?

11

Start Harvest

Harvest progress

Operator	Daniel Sifton
Environment	staging
Mode	Normal
First Run at	April 11, 2017 18:54:52
End Time	April 11, 2017 18:54:57
Duration	5 secs
Records/sec	2.20
Records harvested	11
Records posted to API	11

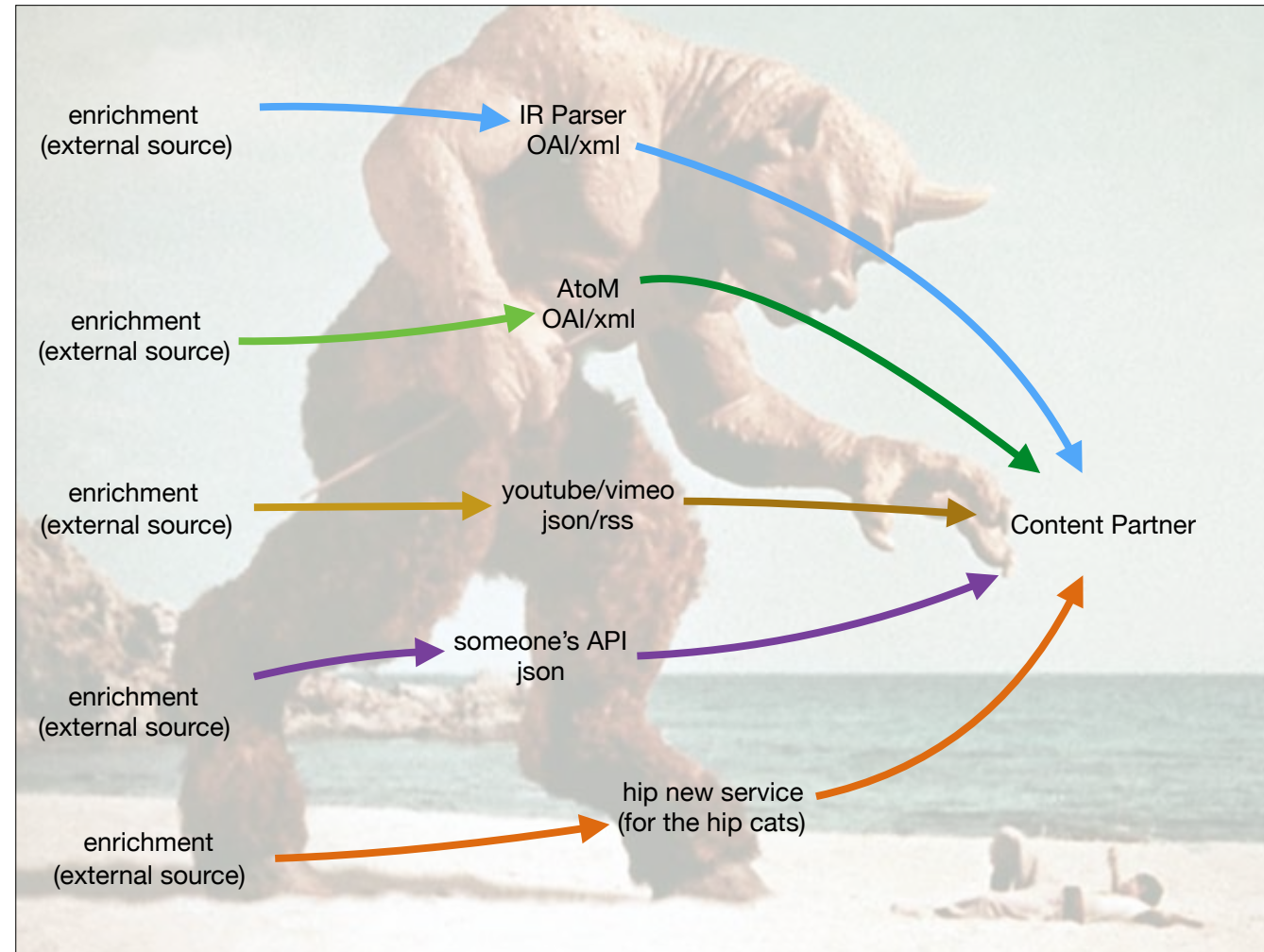
Validation errors

There are no validation errors!

Other errors

There are no other errors.

So, when you're ready, you run the harvest, watch it live! and get a report of your activities. No errors!



When we write a parser and harvest data - our ultimate goal is to bring together different institutions (content partners) through a single discovery layer, right?

We can write parsers that harvest data from different platforms that are used by a single content partner. Those parsers or may not also incorporate metadata from other sources (enrichments)

So, if your institution has a variety of platforms, we harvest from all of them, enrich metadata as required and then pull it all together under a single banner, and make it available alongside dozens of other content partners (who might also have multiple platforms for digital content)

in a sense we have - aggregators within aggregators -  
this is my visual metaphor for challenge and adventure - 7th voyage of Sinbad 1958

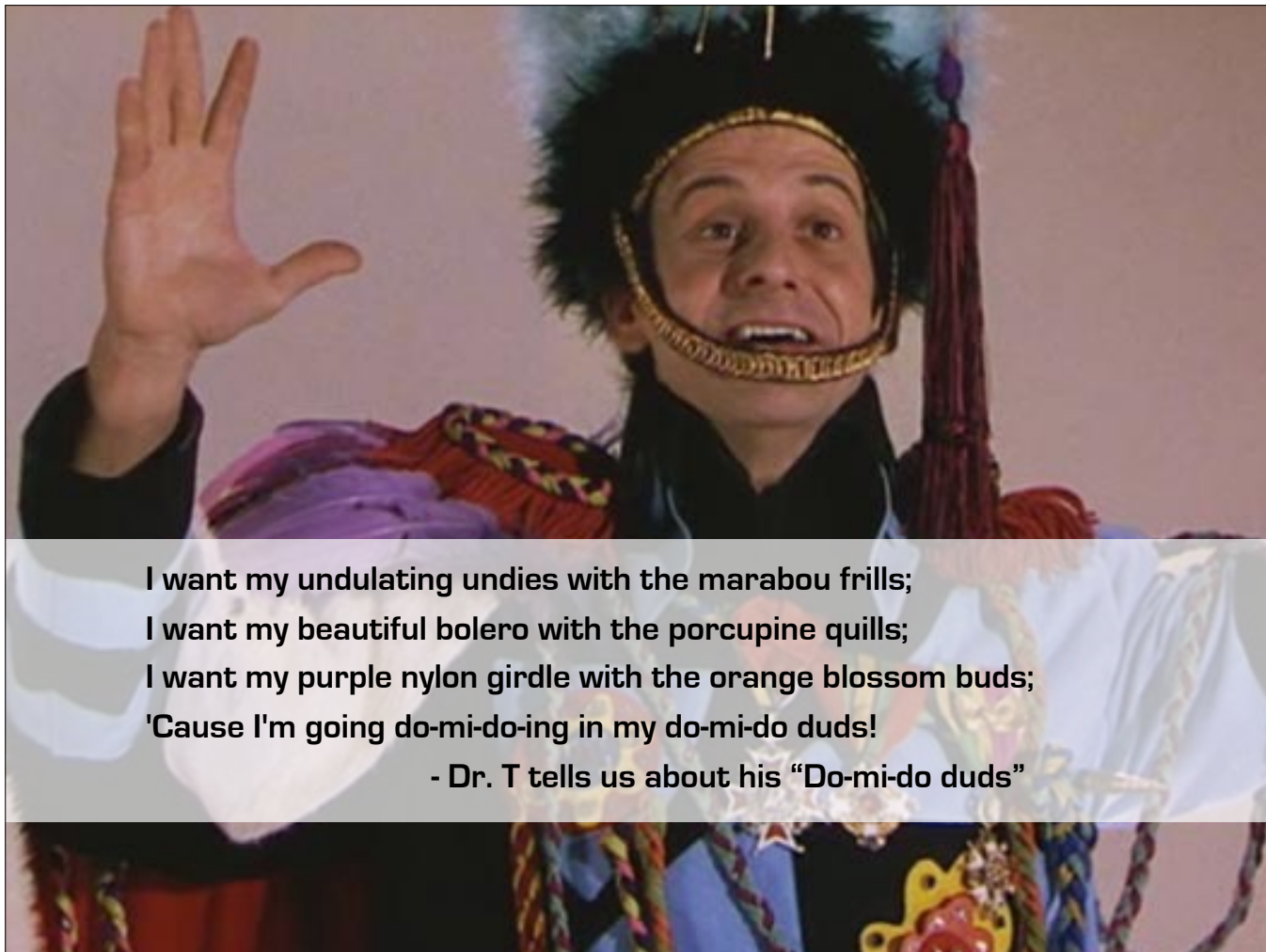
```
t
{
  "record_id": 3199,
  "title": "Salmon Are Sacred - Fraser River Paddle - Chief Marilyn Baptiste",
  "description": "The migration officially launched at a \"Hope for Wild Salmon\" event on 19th October in Hope with other events en route: Chilliwack (20th),
  "creator": {},
  "thumbnail_url": "https://i.ytimg.com/vi/CtmnFjf_Yjw/mqdefault.jpg",
  "locations": []
},
{
  "record_id": 3200,
  "title": "UBCIC AGA 2009 Grand Chief Rose Charlie - The Birth of the UBCIC",
  "description": "September 16, 2009. Grand Chief Rose Charlie speaking about the birth of the Union of BC Indian Chiefs in 1969.",
  "creator": {},
  "thumbnail_url": "https://i.ytimg.com/vi/SS5LpSedQ1c/mqdefault.jpg",
  "locations": []
},
{
  "record_id": 3201,
  "title": "Georgina Papin Memorial - Women's Warrior Song",
  "description": "Missing Women: \"Bringing Our Sisters Home\" Vigil for Georgina Papin, CRAB Park Downtown East Side - September 21, 2010.",
  "creator": {},
  "thumbnail_url": "https://i.ytimg.com/vi/1lJRg_i070I/mqdefault.jpg",
  "locations": []
},
{
  "record_id": 3202,
  "title": "Musqueam People are raising and teaching their children to a Musqueam and",
  "description": "Musqueam People are raising and teaching their children to a Musqueam and",
  "creator": {},
  "thumbnail_url": "https://i.ytimg.com/vi/0pi6N/mqdefault.jpg",
  "locations": []
},
{
  "record_id": 3203,
  "title": "UBCIC AGA 2009 Chief Don Moses - First Chiefs Conference in 1969",
  "description": "September 18, 2010 Chief Don Moses speaking about the first Chiefs conference when the Union of BC Indian Chiefs was formed in 1969.",
  "creator": {},
  "thumbnail_url": "https://i.ytimg.com/vi/dyNG0pASkJU/mqdefault.jpg",
  "locations": []
},
{
  "record_id": 3204,
  "title": "UBCIC 44th AGA. Grand Chief Stewart Phillip - Summary of Recent Events Impacting First Nations",
  "description": "Grand Chief Stewart Phillip, President of the UBCIC, speaking to the UBCIC's 44th AGA.",
  "creator": {},
  "thumbnail_url": "https://i.ytimg.com/vi/DcsZf0xYQyg/mqdefault.jpg",
  "locations": []
},
{
  "record_id": 3205,
  "title": "UBCIC 44th AGA. Jody Woods - Specific Claims and Justice At Last Initiative",
  "description": "UBCIC Research Director Jody Woods updating UBCIC's 44th AGA on specific claims policy and referring to the Justice At Last initiative.",
  "creator": {},
  "thumbnail_url": "https://i.ytimg.com/vi/dFc1B92QYHQ/mqdefault.jpg",
  "locations": []
}
```

Huzzah!

When your harvest is complete this is what you get!

The JSON API results, exciting, no?

you need a front end, a discovery layer to make this palatable for humans, so we built one



**I want my undulating undies with the marabou frills;  
I want my beautiful bolero with the porcupine quills;  
I want my purple nylon girdle with the orange blossom buds;  
'Cause I'm going do-mi-do-ing in my do-mi-do duds!**  
**- Dr. T tells us about his “Do-mi-do duds”**

Have you seen 5000 fingers of Dr. T.?

In the film (the only film Dr. Suess wrote. Dr. T. is the resplendent, a utocratic schoolmaster, who forces children to play his giant piano.

And he really likes fancy dress, maybe at the expense of immediate practicality or possibility. - This prototype front-end is not nearly so refined.

It does not have “marabou frills” it does not have “orange blossom buds”, it illustrates possibilities, not perfection, and I think it does that really really well.

Ready?