

NewspaperWorks:

It's How Samvera Does Newspapers

https://github.com/samvera-labs/newspaper_works

NewspaperWorks is a gem

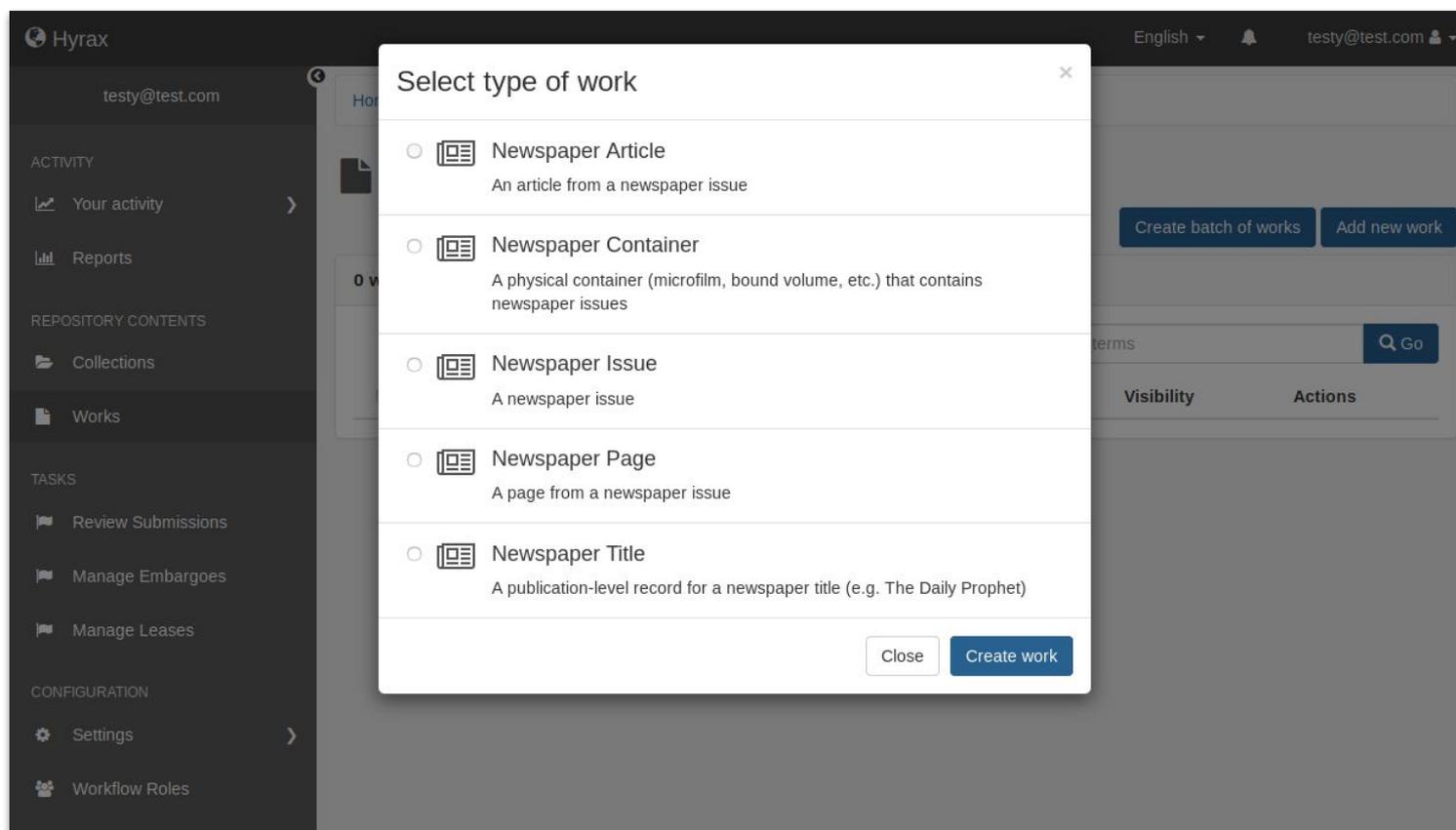
Provides content models, batch ingest workflows, and front-end interaction for digitized newspaper content

- Works with any Hyrax 2.5.*-based repository app
- Easy to install and configure

```
$ vim Gemfile
gem 'newspaper_works'
$ bundle install
$ rails generate newspaper_works:install
```

Newspaper work types & relationships

Title, Container, Issue, Page, Article



The screenshot displays the Hyrax user interface. A dark sidebar on the left contains navigation menus for 'ACTIVITY', 'REPOSITORY CONTENTS', 'TASKS', and 'CONFIGURATION'. The main content area is partially obscured by a white modal dialog box titled 'Select type of work'. The dialog lists five work types, each with a radio button and a brief description. At the bottom of the dialog are 'Close' and 'Create work' buttons. The background interface shows a top navigation bar with 'English' and 'testy@test.com', and a main area with buttons for 'Create batch of works' and 'Add new work'.

testy@test.com

English testy@test.com

testy@test.com

ACTIVITY

- Your activity
- Reports

REPOSITORY CONTENTS

- Collections
- Works

TASKS

- Review Submissions
- Manage Embargoes
- Manage Leases

CONFIGURATION

- Settings
- Workflow Roles

0 v

terms

Go

Visibility Actions

Create batch of works Add new work

Select type of work

- Newspaper Article
An article from a newspaper issue
- Newspaper Container
A physical container (microfilm, bound volume, etc.) that contains newspaper issues
- Newspaper Issue
A newspaper issue
- Newspaper Page
A page from a newspaper issue
- Newspaper Title
A publication-level record for a newspaper title (e.g. The Daily Prophet)

Close Create work

Newspaper metadata fields

Publisher, place of publication, ISSN, LCCN, edition, volume, issue, section, page, geographic coverage, text direction, etc.

The screenshot shows the Hyrax metadata editor interface. At the top, the Hyrax logo is on the left, and the language is set to English, with a user profile for testy@test.com on the right. The main content area is divided into several sections:

- Place of publication:** A dropdown menu is open, showing a search for "Salt". The dropdown list includes: "Salt Lake City, Utah, United States", "Salt, Balqa, Jordan", "Salt Lake City International Airport, Utah, United States", "West Jordan, Utah, United States" (highlighted in blue), "Salt Cay Airport, , Turks and Caicos Islands", "South Jordan, Utah, United States", and "West Valley City, Utah, United States". A "Remove" button is visible to the right of the dropdown.
- OCLC #:** An empty text input field.
- Alternative title:** A text input field with a placeholder "An additional title or subtitle for the item." and a link to "Add another Alternative title".
- Volume:** An empty text input field.
- Edition number:** A text input field with a placeholder "The number of the edition of the item (should be numeric, use '1' as default)."
- Requirements:** A section with two items: "Describe your work" and "Check deposit agreement", both with red error icons.
- Visibility:** A section with radio button options: "Public" (selected), "Institution", "Embargo", "Lease", and "Private". Each option has a brief description of its effect. Below these options is a checkbox for "I have read and agree to the Deposit Agreement" and a "Save" button.

Batch ingest workflows

- National Digital Newspaper Program (NDNP) spec
 - with or without TIFF files
- PDF issues
 - splits PDF into component pages
- TIFF or JP2 page images

```
$ rake newspaper_works:ingest_ndnp -- \
  --path=/path/to/your/ndnp_batch \
  --admin_set=admin_set/default \
  --depositor=admin_user@example.com \
  --visibility=open
```

Batch ingest in action: NDNP

```
vagrant@samvera-newspapers: ~/projects/newspaper_works/internal_test_app (ssh)
vagrant@samvera-newspapers: ~/projects/newspaper_works/internal_test_app (ssh)
vagrant@samvera-newspapers:~/internal_test_app$ bundle exec rake newspaper_works:ingest_ndnp -- \
> --path=/home/vagrant/data/ndnp/batch_test_ver01
2019-10-22 15:43:19 WARN Selenium [DEPRECATION] Selenium::WebDriver::Chrome#driver_path= is deprecated. Use Se
lenium::WebDriver::Chrome::Service#driver_path= instead.
Beginning NDNP batch ingest...
█

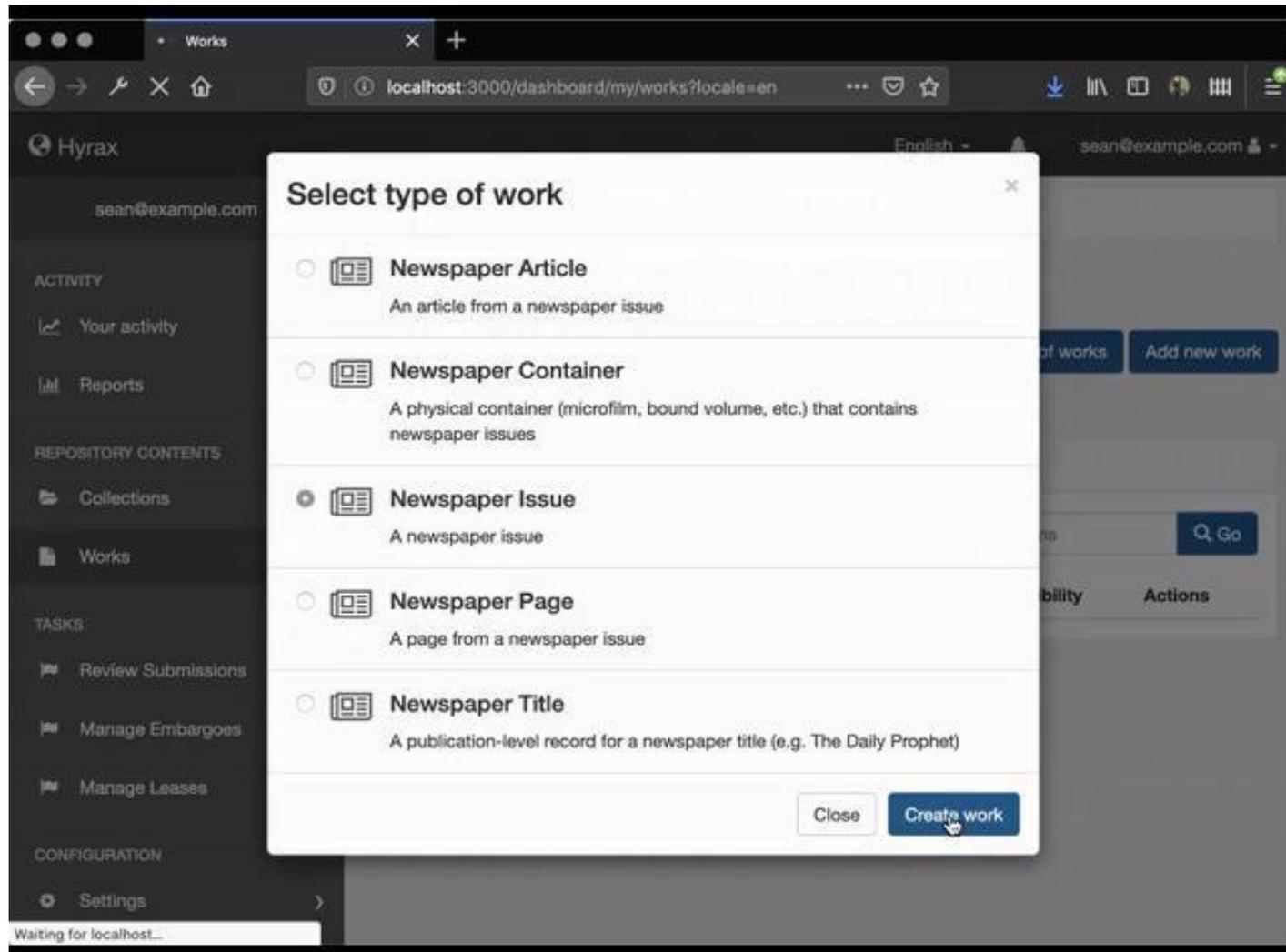
vagrant@samvera-newspapers: ~ (ssh)
}, {"value":6372770976,"max":6396910260}, {"value":6396910260,"max":6421049544}, {"value":6421049544,"max":64451888
28}, {"value":6445188828,"max":6469328112}, {"value":6469328112,"max":6493467396}, {"value":6493467396,"max":651760
6680}, {"value":6517606680,"max":6541745964}, {"value":6541745964,"max":6565885248}, {"value":6565885248,"max":6590
024532}, {"value":6590024532,"max":6614163816}, {"value":6614163816,"max":6638303100}, {"value":6638303100,"max":66
62442384}, {"value":6662442384,"max":6686581668}, {"value":6686581668,"max":6710720952}, {"value":6710720952,"max":6
734860236}, {"value":6734860236,"max":6758999520}, {"value":6758999520,"max":6783138804}, {"value":6783138804,"max"
":6807278088}, {"value":6807278088,"max":6831417372}, {"value":6831417372,"max":6855556656}, {"value":6855556656,"ma
x":6879695940}, {"value":6879695940,"max":6903835224}, {"value":6903835224,"max":6927974508}, {"value":6927974508,"
max":6952113792}, {"value":6952113792,"max":6976253076}, {"value":6976253076,"max":7000392360}, {"value":7000392360,
"max":7024531644}, {"value":7024531644,"max":7048670928}, {"value":7048670928,"max":7072810000}]", [{"seq", 1}, {"ra
nd", "<2537 bytes of binary data>"}, {"updated_at", "2019-10-22 15:43:28.756935"}, {"id", 1}]
2019-10-22 15:43:28 +0000: (3.4ms) commit transaction
█
```

Batch ingest in action: TIFF

```
vagrant@samvera-newspapers: ~/projects/newspaper_works/internal_test_app (ssh)
vagrant@samvera-newspapers: ~/projects/newspaper_works/internal_test_app (ssh)
vagrant@samvera-newspapers:~/internal_test_app$ bundle exec rake newspaper_works:ingest_issues -- --path=/home/vagrant/data/tiff_batch/sn93059126
2019-10-22 15:50:12 WARN Selenium [DEPRECATION] Selenium::WebDriver::Chrome#driver_path= is deprecated. Use Selenium::WebDriver::Chrome::Service#driver_path= instead.
Beginning batch ingest of issues for single publication...
Ingest of issue(s) ingest complete, but may be pending background jobs. See log/ingest.log for details.

vagrant@samvera-newspapers: ~ (ssh)
2019-10-22 15:54:50 +0000: Enqueued FileSetAttachedEventJob (Job ID: 852416d0-306c-4e7a-a3f9-8c71f9af86a0) to Sidekiq(default) with arguments: #<GlobalID:0x000000000e96de40 @uri=#<URI::GID gid://internal/FileSet/h989r3203>>, #<GlobalID:0x000000000e96d210 @uri=#<URI::GID gid://internal/User/1>>
2019-10-22 15:54:50 +0000: Enqueued InheritPermissionsJob (Job ID: 3e7b08d0-f731-416e-a2be-fd09d8a7d2f4) to Sidekiq(default) with arguments: #<GlobalID:0x000000000e4ddfb0 @uri=#<URI::GID gid://internal/NewspaperPage/47429912h>>
2019-10-22 15:54:50 +0000: Enqueued IngestLocalFileJob (Job ID: 4bb4b343-2dbc-4f4a-af68-441f3d05d1df) to Sidekiq(default) with arguments: #<GlobalID:0x000000000dc90510 @uri=#<URI::GID gid://internal/FileSet/h989r3203>>, "/home/vagrant/data/tiff_batch/sn93059126/1853060401/page4.tiff", #<GlobalID:0x000000000dc8f890 @uri=#<URI::GID gid://internal/User/1>>
2019-10-22 15:54:50 +0000: Enqueued NewspaperWorks::ComposeIssuePDFJob (Job ID: 60f911e6-d819-41da-8bb3-bf1b0e24d220) to Sidekiq(default) with arguments: #<GlobalID:0x000000000dc82d70 @uri=#<URI::GID gid://internal/NewspaperIssue/73666448x>>
2019-10-22 15:54:50 +0000: (NewspaperWorks::Ingest::BatchIssueIngester) Issue ingest completed for LCCN sn93059126. Asynchronous jobs may still be creating derivatives for issue, and child page works.
```

Single item ingest: PDF issue



Batch ingest actions

Command-line interface (rake task)

- Parses directories and files
- Creates repository objects
 - Title, Issue, Page
- Adds available metadata
 - publication title, date, place of publication
- Attaches files and creates derivatives
 - PDF, JP2, ALTO XML, JSON word-coordinates
- Can specify AdminSet, depositor, and visibility

Calendar issue browsing

Newspaper Title | Daily national Democrat | ID: 4q77fr32b | Hyrax - Mozilla Firefox

Newspaper Title | Daily national Democrat | ID: 4q77fr32b | Hyrax - Mozilla Firefox

localhost:3000/concern/newspaper_titles/4q77fr32b?locale=en

Issues: 1858 >>

January

Su	Mo	Tu	We	Th	Fr	Sa
27	28	29	30	31	1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31	1	2	3	4	5	6

February

Su	Mo	Tu	We	Th	Fr	Sa
31	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	1	2	3	4	5	6

March

Su	Mo	Tu	We	Th	Fr	Sa
28	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31	1	2	3

April

Su	Mo	Tu	We	Th	Fr	Sa
28	29	30	31	1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	1

May

Su	Mo	Tu	We	Th	Fr	Sa
25	26	27	28	29	30	1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31	1	2	3	4	5

June

Su	Mo	Tu	We	Th	Fr	Sa
30	31	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	1	2	3

July

Su	Mo	Tu	We	Th	Fr	Sa
27	28	29	30	1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

August

Su	Mo	Tu	We	Th	Fr	Sa
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31	1	2	3	4

September

Su	Mo	Tu	We	Th	Fr	Sa
29	30	31	1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	1	2

October

Su	Mo	Tu	We	Th	Fr	Sa
26	27	28	29	30	1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31	1	2	3	4	5	6

November

Su	Mo	Tu	We	Th	Fr	Sa
31	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	1	2	3	4

December

Su	Mo	Tu	We	Th	Fr	Sa
28	29	30	1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	1

Newspapers advanced search

 Hyrax English ▾  Login

Home About Help Contact Search Hyrax

Search Newspapers

Use this form to search for full-text newspaper content.

Keywords

Limit search to front pages:

Date range

Publication title

- Daily national Democrat
- The Park Record
- Weekly Trinity journal

Language

- English

Newspapers search tips:

- For an open-ended date range limit (e.g. before 1850, after 1900, etc.) leave either the "Start" or "End" field blank in the "Date Range" section.
- If a date range is entered in the form in the "Date Range" section, undated items will not be included in the search results.
- If no boxes are checked in the field limit section, all values will be included.
- Use quotation marks to search as a phrase.
- Use "OR", "AND", and "NOT" within the keyword field to create complex boolean logic. You can also use parentheses.
- Use "+" before a term to make it required.
- Use "-" before a word or phrase to exclude it.

Search within title; browse front pages

 Newspaper Title

The Park Record Public

Search 

Limit search to front pages:

View all front pages

Keyword matches in search results

Hyrax English testy@test.com

Home About Help Contact Search Hyrax president Go All

Limit your search

- Place of publication
- Publication title
- Type
- Resource type
- Language

Filtering by: president Start Over

1 - 2 of 2 Sort by publication date 10 per page

[Daily national Democrat: August 13, 1858: Page 2](#)



Publication: Daily national Democrat
Publication date: 08/13/1858
Place of publication: Marysville, California, United States
Keyword matches: ... **President**. wui the other will certainly slay the **President***,* rty itl t j,j* State. Our readers are aware ...
more >>
Owner: batchuser@example.com
Date Uploaded: 10/17/2019
Date Modified: 10/17/2019
Resource Type: Newspapers
Identifier: 0007

[Weekly Trinity journal: February 14, 1857: Page 1](#)



Publication: Weekly Trinity journal
Publication date: 02/14/1857
Place of publication: Weaverville Joss House State Historic Park, California, United States
Keyword matches: ... the head of the Medical Faculty in London, and twice **President** of the ! London College of

Full text search in viewer

Daily national Democrat / August 13, 1858

Image 1 of 2 Go

CONTENTS

Daily national Democrat: August 13, 1858

MORE INFORMATION

18ff7159... b38f910...

PRINTING OFFICE,
and Street and Maiden Lane,
MARYSVILLE.
of the above named Establishment
y announce to the citizens of Marys-
ounding country, that they are now
ate all kinds of
and Ornamental Job
PRINTING
URPASSED by any other Establish-
a.
PRINTING PRESSES, of the latest

ly engaged in daming (*damning*) the
rivers. Such appears to be the voca-
tion of all who have returned from
those regions, so far as we can learn.—
Sonoma County Journal.

WASN'T I THERE TOO.—At a recent
election in this State, a lad presented
himself at the poll to claim the benefit
of the elective franchise. —
Feeling a deep interest in a favorite
candidate, the father, who was evident-
ly opposed to the boys preference, stood
at the ballot box, and challenged his
right to vote, on the ground of his not
being of age. The young man declared
that he was twenty-one years old, that

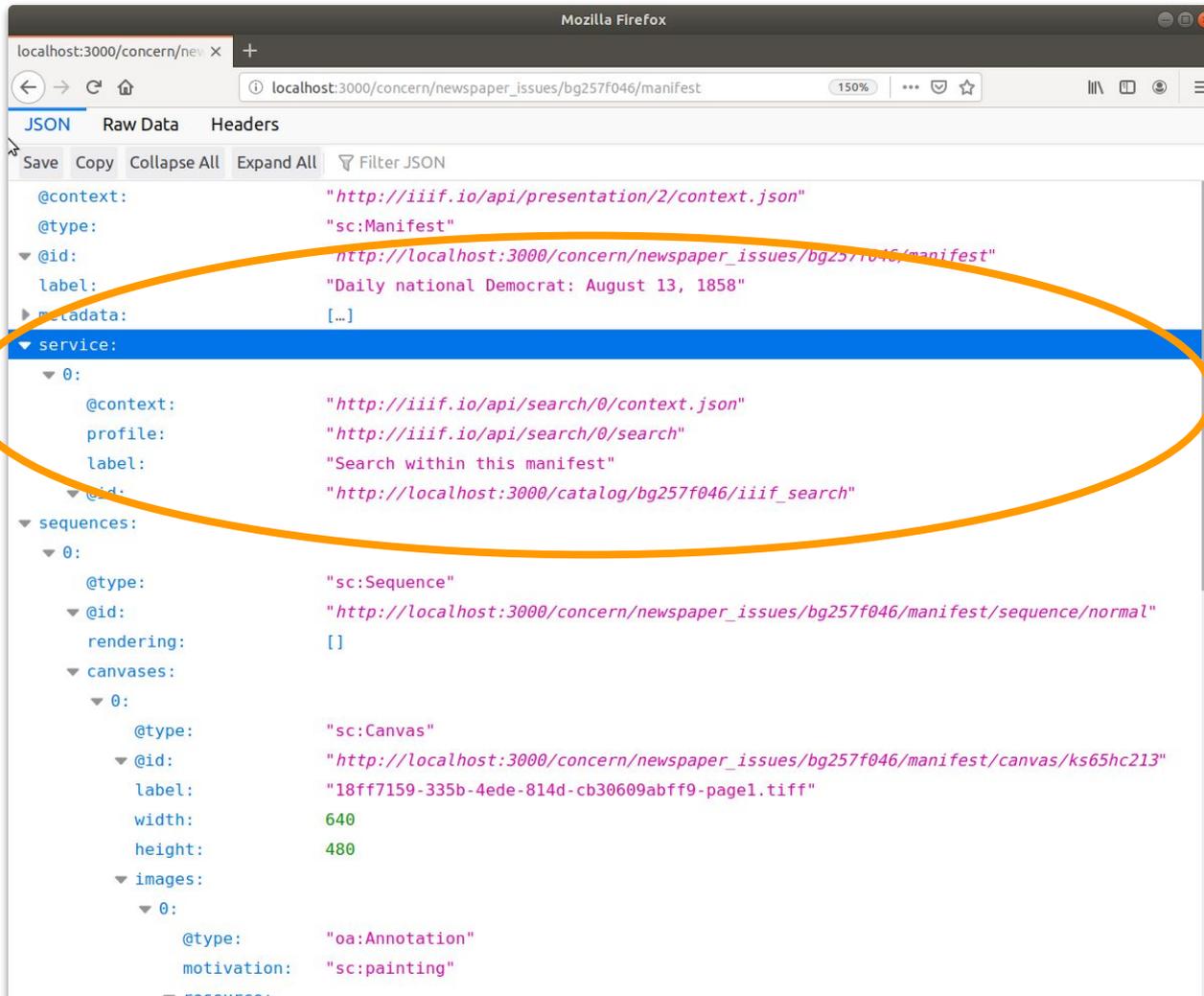
Miscellaneous.
NEW YORK
Livery & Sale Stable
FORMERLY MONTGOMERY & JOHNSON'S.
FIRE PROOF BRICK BUILDING.
Maiden Lane,
Between 1st & 2d sts.,
Carriages, Buggies, and good Saddle Horses kept
constantly on hand, for hire or sale, on moderate
terms.
Particular attention paid to Horses on Livery.
FIELDING & CO.
New Administration.
LEWIS, THE LAMP SALT

M. R. P. K.
partner
ter he condac
& O'FARRAL.
Marysville,
OSGOOD
Flour, G
CL
First street
no26ay2p
JNO. M. MAT

page 18ff7159-335b-4ede-814d-cb30609abff9-page1.tiff of

Search within this item:
election

IIIF: Presentation manifest + Content Search



The screenshot displays a Mozilla Firefox browser window with the address bar showing `localhost:3000/concern/newspaper_issues/bg257f046/manifest`. The browser's developer tools are open, showing the JSON response of the manifest. The JSON is structured as follows:

```
{
  "@context": "http://iiif.io/api/presentation/2/context.json",
  "@type": "sc:Manifest",
  "@id": "http://localhost:3000/concern/newspaper_issues/bg257f046/manifest",
  "label": "Daily national Democrat: August 13, 1858",
  "metadata": [...],
  "service": {
    "0": {
      "@context": "http://iiif.io/api/search/0/context.json",
      "profile": "http://iiif.io/api/search/0/search",
      "label": "Search within this manifest",
      "@id": "http://localhost:3000/catalog/bg257f046/iiif_search"
    }
  },
  "sequences": {
    "0": {
      "@type": "sc:Sequence",
      "@id": "http://localhost:3000/concern/newspaper_issues/bg257f046/manifest/sequence/normal",
      "rendering": [],
      "canvases": {
        "0": {
          "@type": "sc:Canvas",
          "@id": "http://localhost:3000/concern/newspaper_issues/bg257f046/manifest/canvas/ks65hc213",
          "label": "18ff7159-335b-4ede-814d-cb30609abff9-page1.tiff",
          "width": 640,
          "height": 480,
          "images": {
            "0": {
              "@type": "oa:Annotation",
              "motivation": "sc:painting"
            }
          }
        }
      }
    }
  }
}
```

The `service` section is highlighted with a blue background and an orange oval. The `sequences` section contains a `sequence/normal` which includes a `sequence/normal` and a `sequence/normal` canvas with a `sequence/normal` image.

Blacklight IIIF Search

Gem that provides an endpoint in your Blacklight app that will return a JSON response conforming to IIIF Content Search API v. 1.0.

Integrate the URL for this service into IIIF Presentation manifests to allow clients/viewers to provide functionality for searching within a resource and displaying results.

```
$ vim Gemfile
gem 'blacklight_iiif_search'
$ bundle install
$ rails generate blacklight_iiif_search:install
```

Hierarchical navigation

HyraX English testy@test.com

Home About Help Contact Search HyraX Enter search terms Go All

Home / Dashboard / Works / Daily national Democrat:...

Newspaper Page

Daily national Democrat: August 13, 1858: Page 2

Public Deposited

Edit Delete Attach Child Add to collection Feature

Daily national Democrat / August 13, 1858 / Page 2 << Previous | Next

Daily national Democrat: August 13, 1858: Page 2

>> MORE INFORMATION

About the item

Title
Daily national Democrat: August 13, 1858: Page 2

Page number
2



Semantic URLs

Predictable URL based on LCCN, pub date, edition, page

Publication date

08/13/1858

Identifier

0007

Resource type

Newspapers

Persistent url

<http://localhost/newspapers/sn84038814/1858-08-13/ed-1/seq-2>

Features in development

- Batch ingest of article-segmented content
- Download clipped image segment
- Geographic browsing & visualization
- Timeline browsing & visualization
- Hyrax 3.0 compatibility

How to get started

Wiki:

- Detailed docs on installation, setup, object modeling, metadata profile, batch ingest, and more

https://github.com/samvera-labs/newspaper_works/wiki

Test site:

- Hyrax repository application running latest release of gem

<https://newspaperworks.digitalnewspapers.org/>

Vagrant box:

- Installs all dependencies (including Fedora and Solr)
- Creates test app with NewspaperWorks installed

<https://github.com/marriott-library/samvera-newspapers-vagrant>

NewspaperWorks needs you!

Samvera Newspapers Interest Group:

- Meets bi-monthly on 1st Thursday 1 PM ET/11AM MT

<https://wiki.duraspace.org/display/samvera/Samvera+Newspapers+Interest+Group>

Samvera Slack:

- #newspapers channel

<https://samvera.slack.com>

Acknowledgements

- Institute of Museum & Library Services
- Grant advisory committee
 - Michael Appleby
 - Jenn Colt
 - Ryan Cordell
 - Michael Della Bitta
 - Karen Estlund
 - Tom Ewing
 - Michael Giarlo
 - John Herbert
- Newspapers Interest Group

Questions? Contact us

```
|-----|-----|  
| Eben English | eenglish [at] bpl [dot] org  
| Brian McBride | brian.mcbride [at] utah [dot] edu  
| Jacob Reed | jacob.reed [at] utah [dot] edu  
| Sean Upton | sean.upton [at] utah [dot] edu
```