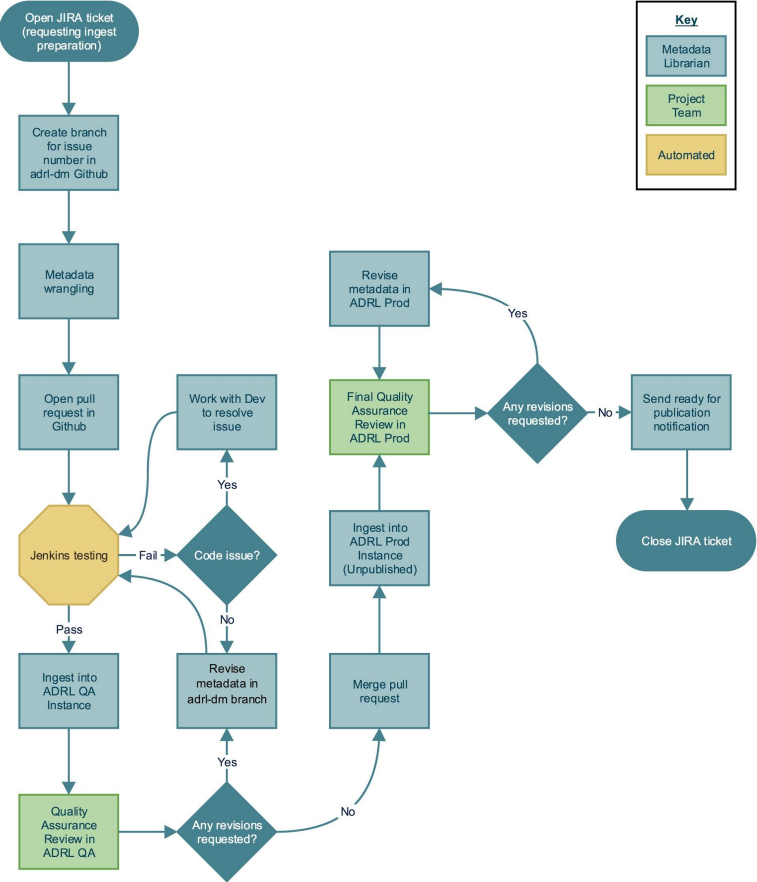


Metadata Management and Batch Ingest Automation

With Git/GitHub, Jenkins, and JIRA

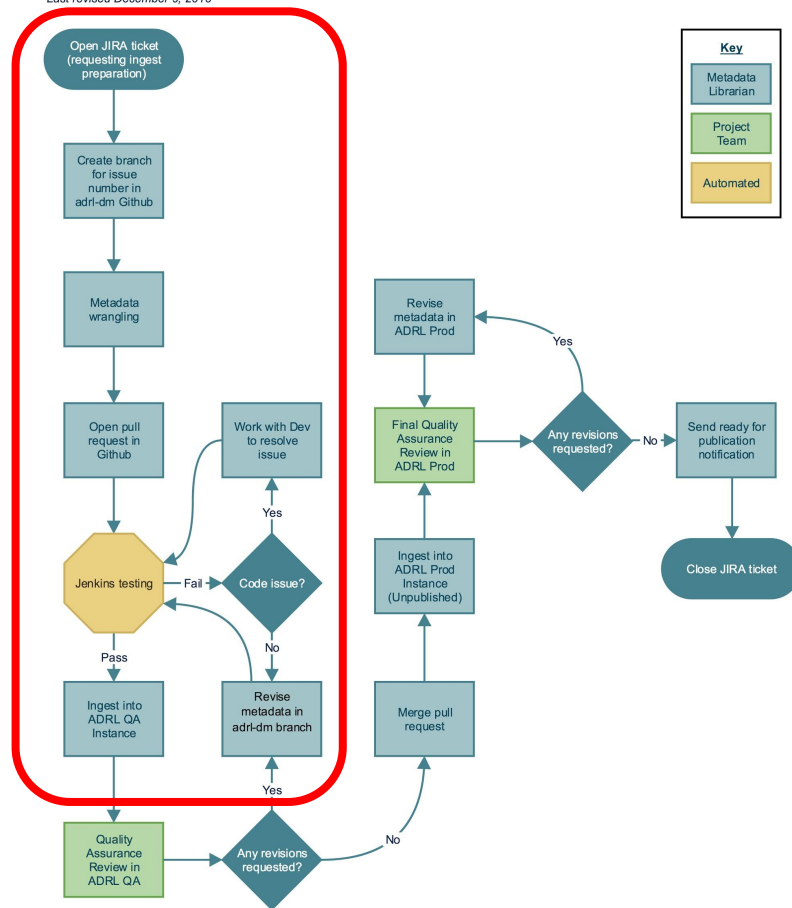
Alex Dunn (UC Santa Barbara)
Chrissy Rissmeyer (UC Santa Barbara)

ADRL Ingest Preparation Workflow
(Non-Harvest Objects)
Last revised December 9, 2016



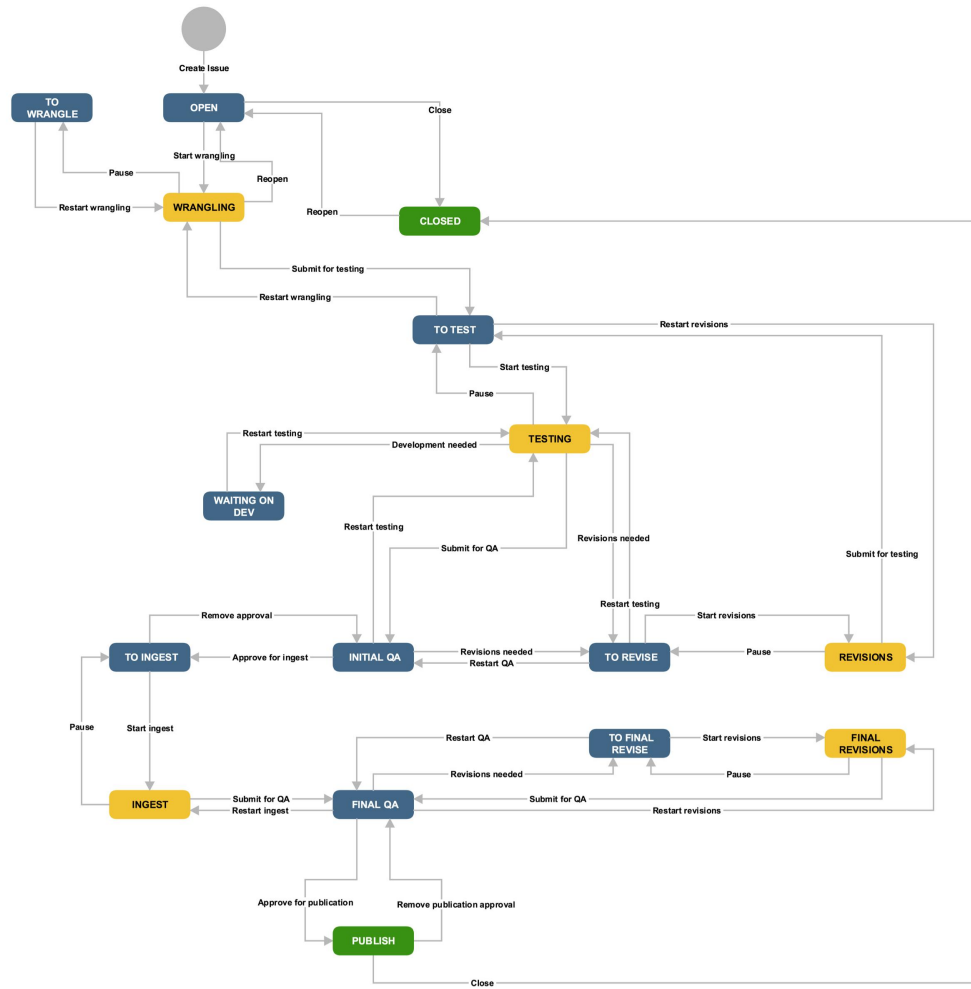
ADRL Ingest Preparation Workflow (Non-Harvest Objects)

Last revised December 9, 2016



Open JIRA ticket
(requesting ingest
preparation)







Edit



Comment



Assign



More ▾



Reopen



Back to Publish



Close



Admin ▾



Share



Export ▾

Export ▾

Details

Type: ☒ General Ingest Status: **CLOSED**

Priority: Normal Resolution: Done

Component/s: [CSV Input Stream](#)

Labels: None

Project ID: mss228

Content Files: [svmwindows.library.ucsb.edu/repository/images/mss228](#)

GitHub Pull Request: <https://github.library.ucsb.edu/CMS/adrl-dm/pull/2>

Initial QA Link: <http://adrlstage.library.ucsb.edu/collections/fk4zw1qj9f>

Final QA Link: <http://alexandria.ucsb.edu/collections/fk4ww7nt8p>

Number of Objects: 1,557

Number of Files: 1,577 TIFFs

Description

1557 TIFF files (1557 objects)

Attachments



Drop files to attach, or [browse](#).



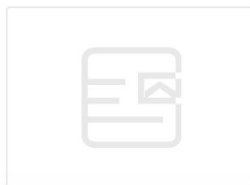
[batch_ingest_final_qa_chec](#)

03/23/2017 10:10 PM 500 KB



[batch_ingest_final_qa_chec](#)

03/23/2017 10:11 PM 500 KB



[batch_ingest_initial_qa_che](#)

03/23/2017 10:00 AM 500 KB

People

Assignee: Chrissy Rissmeyer

Reporter: Chrissy Rissmeyer

Project Lead: Chrissy Rissmeyer

Metadata Analyst: Chrissy Rissmeyer

Quality Assurance: Chrissy Rissmeyer

Watchers: [Stop watching this issue](#)

Dates

Created: 01/30/2017 11:58 AM

Updated: 05/08/2017 01:53 PM

Resolved: 03/21/2017 12:21 PM

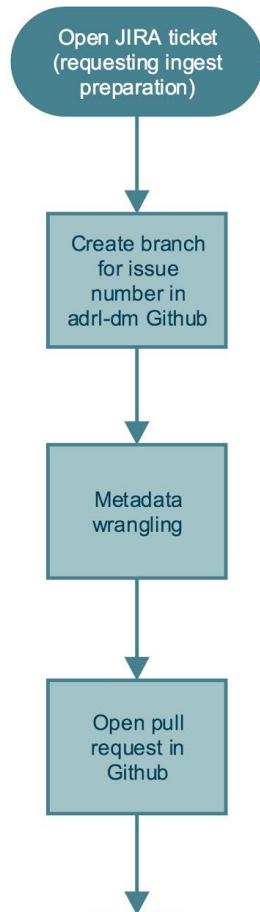
Requested Due Date: 04/11/2017

Development

[Create branch](#)

Agile

[View on Board](#)





This repository

Search

Pull requests

Issues

Gist



CMS / adrl-dm

Watch

4

Star

2

Fork

1

Code

Issues 0

Pull requests 11

Wiki

Pulse

Graphs

Settings

No description or website provided. — Edit

224 commits

17 branches

0 releases

2 contributors

Branch: master

New pull request

Create new file

Upload files

Find file

Clone or download

crissmeyer committed on GitHub Enterprise Merge pull request #26 from CMS/jenkins-filePath

Latest commit a46c671 on Jul 13

archive	converts to csv file and archives excel	8 months ago
in-process	moves csv to ingest ready folder	8 months ago
ingest-ready	adds file path to hpamss032	4 months ago
templates	Updated end_qualifier date column headers to finish_qualifer (created...	2 years ago

General Rules

- **Rule #1:** Don't ingest content into the production instance of ADRL before the code is ready for it! (Unless you really, really have to.)
- **Rule #2:** All metadata changes should be made on a branch first. A pull request (PR) to merge with the master branch should be made when the objects are **ready** and **able** to be ingested into the production instance of ADRL . Move metadata into `ingest-ready` directory before opening pull request.
- **Rule #3:** Ingests into the production instance of ADRL may only be done from the master branch.

Branching Guidelines

master = ingest-ready folder contains objects that are ready and able to be ingested into the production instance of ADRL

When to create a branch?

1. New collection that is not ready for ingest (either metadata is in-process or code required for ingest is still under development)
2. Metadata adjustment to previously ingested collection that requires a code change & re-ingest (i.e., rightsstatements.org implementation)
3. Batch metadata updates not requiring a code change (to be done very carefully)
4. Development team needs sample records for their work. This metadata may or may not be "real" metadata. PRs for sample branches should never be made.

How to name branches?

1. New collection that is not ready for ingest = use corresponding ADRL Ingest ticket number (i.e., INGEST-2)
2. Metadata adjustment to previously ingested collection that requires a code change & re-ingest = use corresponding DIGREPO ticket number (i.e., DIGREPO-493)
3. Batch metadata updates = use corresponding ticket number (note that JIRA project for this type of work does not exist yet)
4. Sample records = samples-type (i.e., samples-maps)



This repository Search

Pull requests Issues Gist



CMS / adri-dm

Watch 4 Star 2 Fork 1

Code Issues 0 Pull requests 11 Wiki Pulse Graphs Settings

Overview Yours Active Stale All branches

Search branches...

Default branch

master Updated 4 months ago by crissmeyer

Default

Change default branch

Your branches

INGEST-10 Updated 24 days ago by Chrissy Rissmeyer



#18

Open



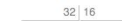
samples-newspapers Updated a month ago by Chrissy Rissmeyer

0 | 2

New pull request



INGEST-18 Updated 4 months ago by crissmeyer



New pull request



INGEST-13 Updated 2 months ago by Tom Brittnacher



#13

Open



INGEST-12 Updated 2 months ago by Tom Brittnacher



#12

Open



View more of your branches >

Active branches

INGEST-10 Updated 24 days ago by Chrissy Rissmeyer



#18

Open



samples-newspapers Updated a month ago by Chrissy Rissmeyer

0 | 2

New pull request



INGEST-13 Updated 2 months ago by Tom Brittnacher



#13

Open



INGEST-12 Updated 2 months ago by Tom Brittnacher



#12

Open



INGEST-8 Updated 2 months ago by Tom Brittnacher



#15

Open



View more active branches >

<> Code

! Issues 0

🔗 Pull requests 11

📖 Wiki

📶 Pulse

📊 Graphs

⚙️ Settings

Branch: INGEST-10 ▾



Commits on Oct 10, 2017

**updates invalid date "#N/A"**

Chrissy Rissmeyer committed 28 days ago



192d6c7

**Merge remote-tracking branch 'origin/INGEST-10' into INGEST-10**

Chrissy Rissmeyer committed 28 days ago



841a9c2

**Revert "updates invalid date "#N/A""** ...

Chrissy Rissmeyer committed 28 days ago



94b2925

**updates invalid date "#N/A"**

Chrissy Rissmeyer committed 28 days ago



073d2bc



Commits on Sep 11, 2017

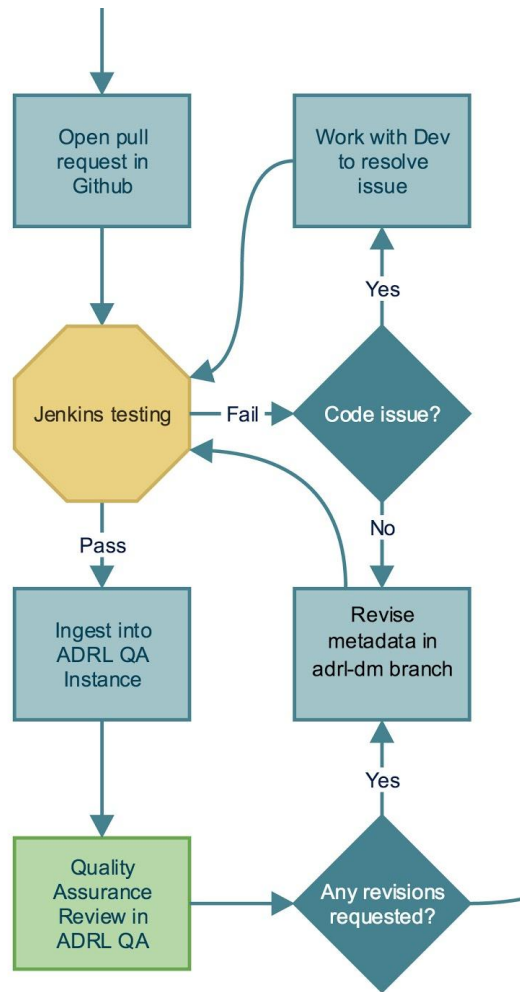
**updates path name to reflect source folder name change**

Tom Brittnacher committed on Sep 11



49ac233





Programmatic Metadata Validation

- <https://github.com/ucsblibrary/metadata-ci>
- Provides modular library and CLI interface
- Add your own custom metadata checks!

Typing `bin/check` is Too Much Work

- Our metadata validation tool can be integrated into CI systems
- Don't merge invalid metadata into `master`!
- <https://github.com/ucsblibrary/metadata-ci/blob/master/bin/jenkins-pr>

A Webhook is a Fancy Word for POST

Webhooks / Manage webhook

We'll send a POST request to the URL below with details of any subscribed events. You can also specify which data format you'd like to receive (JSON, x-www-form-urlencoded, etc). More information can be found in [our developer documentation](#).

Payload URL *

Content type

Secret

***** — [Edit](#)


Which events would you like to trigger this webhook?


☐ Just the push event.





☐ Send me **everything**.


☒ Let me select individual events.

Jenkins (or Travis) sends a POST back to GitHub



**Some checks were not successful**[Hide all checks](#)
1 failing and 1 successful checks

	 validate — Build finished.	Details
	 ingest — Build finished.	Details

**This branch has no conflicts with the base branch**
Only those with [write access](#) to this repository can merge pull requests.

Benefits: Technical & Otherwise

- Lightweight solution
- Repository agnostic
- Less manual work by humans
- Increased autonomy / less developer time
- Increased collaboration & understanding

Future Directions

- More Jenkins tests!
- Automatic deployments
- Possible tool integration

Questions?

Thank you!

Alex Dunn - adunn@ucsb.edu / @alexdunn (Slack)

Chrissy Rissmeyer - crissmeyer@ucsb.edu / @chrissy (Slack)
github.com/ucsblibrary/metadata-ci

Backup Demo Slides

sc2017-demo-test #29



crissmeyer wants to merge 2 commits into master from sc2017-demo-test2

Conversation 0

Commits 2

Files changed 1

+156 -0



crissmeyer commented 7 minutes ago

CMS member



Wonky dates + missing column header



Chrissy Riss... added some commits an hour ago



adds wonky dates

X 39d7fdf



fixes dates and header

● 7b8e8b3

Add more commits by pushing to the sc2017-demo-test2 branch on CMS/adrl-dm.



Some checks haven't completed yet

[Hide all checks](#)

2 pending checks



● ingest — Build started sha1 is merged.

[Details](#)

● validate — Build started sha1 is merged.

[Details](#)

This branch has no conflicts with the base branch

Merging can be performed automatically.



Merge pull request

You can also [open this in GitHub Desktop](#) or view [command line instructions](#).

Labels

None yet

Milestone

No milestone

Assignees

No one—assign yourself

Labels

None yet

Milestone

No milestone

Assignees

No one—assign yourself

1 participant



Notifications

sc2017-demo-test #29

Edit

 **Open** crissmeyer wants to merge 1 commit into master from sc2017-demo-test2 Conversation 0 Commits 1 Files changed 1

+156 -0



crissmeyer commented just now

CMS member



Wonky dates + missing column header



adds wonky dates

39d7fdf

Add more commits by pushing to the **sc2017-demo-test2** branch on **CMS/adri-dm**.

Some checks haven't completed yet

[Hide all checks](#)

1 failing and 1 pending checks

✗ **validate** — Build finished.[Details](#)● **ingest** — Build started sha1 is merged.[Details](#)

This branch has no conflicts with the base branch

Merging can be performed automatically.



Merge pull request

You can also [open this in GitHub Desktop](#) or view [command line instructions](#).

Labels



None yet

Milestone



No milestone

Assignees



No one—assign yourself

Labels



None yet

Milestone



No milestone

Assignees



No one—assign yourself

1 participant



sc2017-demo-test #29

Edit

 **Open** crissmeyer wants to merge 1 commit into master from sc2017-demo-test2

💬 Conversation 0

🔗 Commits 1

📄 Files changed 1

+156 -0



crissmeyer commented just now

CMS member



Wonky dates + missing column header

🔗 adds wonky dates

✗ 39d7fdf

Add more commits by pushing to the **sc2017-demo-test2** branch on **CMS/adrl-dm**.**Some checks were not successful**[Hide all checks](#)

1 failing and 1 successful checks

✗ **validate** — Build finished.[Details](#)✓ **ingest** — Build finished.[Details](#)**This branch has no conflicts with the base branch**

Merging can be performed automatically.

**Merge pull request**You can also [open this in GitHub Desktop](#) or view [command line instructions](#).

Labels



None yet

Milestone



No milestone

Assignees



No one—assign yourself

Labels



None yet

Milestone



No milestone

Assignees



No one—assign yourself

1 participant



Back to Project

Status

Changes

Console Output

View as plain text

View Build Information

Parameters

Environment Variables

Embeddable Build Status

Previous Build

Next Build

Console Output

```
GitHub pull request #29 of commit 39d7fdf9b5af0f6b1467fa15fc3b06611d70d53b, no merge conflicts.
Setting status of 39d7fdf9b5af0f6b1467fa15fc3b06611d70d53b to PENDING with url
http://jenkins.library.ucsb.edu:8080/job/metadata\_pr/51/ and message: 'Build started shal is merged.'
Using context: validate
[EnvInject] - Loading node environment variables.
Building remotely on adrljenkinsbuild1 (RSpec) in workspace /home/jenkins/workspace/metadata_pr
Wiping out workspace first.
Cloning the remote Git repository
Cloning repository https://github.library.ucsb.edu/CMS/adrl-dm.git
> git init /home/jenkins/workspace/metadata_pr # timeout=10
Fetching upstream changes from https://github.library.ucsb.edu/CMS/adrl-dm.git
> git --version # timeout=10
using GIT_ASKPASS to set credentials AD credentials
Setting http proxy: 10.3.100.201:3128
> git fetch --tags --progress https://github.library.ucsb.edu/CMS/adrl-dm.git
+refs/heads/*:refs/remotes/origin/*
> git config remote.origin.url https://github.library.ucsb.edu/CMS/adrl-dm.git # timeout=10
> git config --add remote.origin.fetch +refs/heads/*:refs/remotes/origin/* # timeout=10
> git config remote.origin.url https://github.library.ucsb.edu/CMS/adrl-dm.git # timeout=10
Fetching upstream changes from https://github.library.ucsb.edu/CMS/adrl-dm.git
using GIT_ASKPASS to set credentials AD credentials
Setting http proxy: 10.3.100.201:3128
> git fetch --tags --progress https://github.library.ucsb.edu/CMS/adrl-dm.git
+refs/pull/*:refs/remotes/origin/pr/*
> git rev-parse refs/remotes/origin/pr/29/merge^{commit} # timeout=10
> git rev-parse refs/remotes/origin/pr/29/merge^{commit} # timeout=10
Checking out Revision 8917b87bcdce26e4c6a173a07a129cc5b1fed8bd (refs/remotes/origin/pr/29/merge)
> git config core.sparsecheckout # timeout=10
> git checkout -f 8917b87bcdce26e4c6a173a07a129cc5b1fed8bd
First time build. Skipping changelog.
```

```
Using hamster 3.0.0
Using tzinfo 1.2.3
Using pry 0.10.4
Using parser 2.4.0.0
Using rdf 2.2.9
Using activesupport 5.1.4
Using pry-doc 0.11.1
Using rubocop 0.49.1
Using rdf-vocab 2.2.6
Using nom-xml 0.6.0
Using mods 2.1.0
Using metadata_ci 0.0.1 from source at `.`
Bundle complete! 9 Gemfile dependencies, 35 gems now installed.
Use `bundle info [gemname]` to see where a bundled gem is installed.
Running checks: ControlledVocabularies, Date, Encoding, Entities, Headers, Schema

* (InvalidDate) ../ingest-ready/sc2017-demo/sc2017-demo.csv:
  created_start on line 26: 'N/A' is not W3C-valid (https://www.w3.org/TR/1998/NOTE-datetime-19980827).

* (InvalidDate) ../ingest-ready/sc2017-demo/sc2017-demo.csv:
  created_start on line 42: '1913-31-06' is not W3C-valid (https://www.w3.org/TR/1998/NOTE-datetime-19980827).

* (InvalidDate) ../ingest-ready/sc2017-demo/sc2017-demo.csv:
  created_start on line 110: '1897-13' is not W3C-valid (https://www.w3.org/TR/1998/NOTE-datetime-19980827).

* (InvalidHeader) ../ingest-ready/sc2017-demo/sc2017-demo.csv:
  '' is not an allowed header.

Build step 'Execute shell' marked build as failure
Setting status of 39d7fdf9b5af0f6b1467fa15fc3b06611d70d53b to FAILURE with url
http://jenkins.library.ucsb.edu:8080/job/metadata\_pr/51/ and message: 'Build finished. '
Using context: validate
Finished: FAILURE
```

sc2017-demo-test #29



crissmeyer wants to merge 2 commits into master from sc2017-demo-test2

💬 Conversation 0

📦 Commits 2

📄 Files changed 1

+156 -0



crissmeyer commented 12 minutes ago

CMS member



Wonky dates + missing column header



Chrissy Riss... added some commits an hour ago



adds wonky dates

✗ 39d7fdf



fixes dates and header

✓ 7b8e8b3

Add more commits by pushing to the sc2017-demo-test2 branch on CMS/adrl-dm.



All checks have passed

2 successful checks

[Hide all checks](#)

✓ ingest — Build finished.

[Details](#)

✓ validate — Build finished.

[Details](#)

This branch has no conflicts with the base branch

Merging can be performed automatically.



Merge pull request

You can also [open this in GitHub Desktop](#) or view [command line instructions](#).

Labels

None yet

Milestone

No milestone

Labels

None yet

Milestone

No milestone

Assignees

No one—assign yourself

1 participant



Notifications

🔊 Unsubscribe

You're receiving notifications

[Back to Project](#)[Status](#)[Changes](#)[Console Output](#)[View as plain text](#)[View Build Information](#)[Parameters](#)[Environment Variables](#)[Embeddable Build Status](#)[Previous Build](#)

Console Output

```
GitHub pull request #29 of commit 7b8e8b3994c142921ba32dc29c39f130df338d40, no merge conflicts.
Setting status of 7b8e8b3994c142921ba32dc29c39f130df338d40 to PENDING with url
http://jenkins.library.ucsb.edu:8080/job/metadata\_pr/52/ and message: 'Build started shal is merged.'
Using context: validate
[EnvInject] - Loading node environment variables.
Building remotely on adrljenkinsbuild1 (RSpec) in workspace /home/jenkins/workspace/metadata_pr
Wiping out workspace first.
Cloning the remote Git repository
Cloning repository https://github.library.ucsb.edu/CMS/adrl-dm.git
> git init /home/jenkins/workspace/metadata_pr # timeout=10
Fetching upstream changes from https://github.library.ucsb.edu/CMS/adrl-dm.git
> git --version # timeout=10
using GIT_ASKPASS to set credentials AD credentials
Setting http proxy: 10.3.100.201:3128
> git fetch --tags --progress https://github.library.ucsb.edu/CMS/adrl-dm.git
+refs/heads/*:refs/remotes/origin/*
> git config remote.origin.url https://github.library.ucsb.edu/CMS/adrl-dm.git # timeout=10
> git config --add remote.origin.fetch +refs/heads/*:refs/remotes/origin/* # timeout=10
> git config remote.origin.url https://github.library.ucsb.edu/CMS/adrl-dm.git # timeout=10
Fetching upstream changes from https://github.library.ucsb.edu/CMS/adrl-dm.git
using GIT_ASKPASS to set credentials AD credentials
Setting http proxy: 10.3.100.201:3128
> git fetch --tags --progress https://github.library.ucsb.edu/CMS/adrl-dm.git
+refs/pull/*:refs/remotes/origin/pr/*
> git rev-parse refs/remotes/origin/pr/29/merge^{commit} # timeout=10
> git rev-parse refs/remotes/origin/origin/pr/29/merge^{commit} # timeout=10
Checking out Revision 4e141a59d81e8c7496772c4930bb952ed99ce6b3 (refs/remotes/origin/pr/29/merge)
> git config core.sparsecheckout # timeout=10
> git checkout -f 4e141a59d81e8c7496772c4930bb952ed99ce6b3
First time build. Skipping changelog.
```

```
Using coderay 1.1.2
Using slop 3.6.0
Using yard 0.9.9
Using parallel 1.12.0
Using ast 2.3.0
Using powerpack 0.1.1
Using ruby-progressbar 1.8.3
Using unicode-display_width 1.3.0
Using trollop 2.1.2
Using bundler 1.15.1
Using rainbow 2.2.2
Using nokogiri 1.8.0
Using hamster 3.0.0
Using tzinfo 1.2.3
Using pry 0.10.4
Using parser 2.4.0.0
Using rdf 2.2.9
Using activesupport 5.1.4
Using pry-doc 0.11.1
Using rubocop 0.49.1
Using rdf-vocab 2.2.6
Using nom-xml 0.6.0
Using mods 2.1.0
Using metadata_ci 0.0.1 from source at `.`
Bundle complete! 9 Gemfile dependencies, 35 gems now installed.
Use `bundle info [gemname]` to see where a bundled gem is installed.
Running checks: ControlledVocabularies, Date, Encoding, Entities, Headers, Schema

No errors found!
Setting status of 7b8e8b3994c142921ba32dc29c39f130df338d40 to SUCCESS with url
http://jenkins.library.ucsb.edu:8080/job/metadata\_pr/52/ and message: 'Build finished. '
Using context: validate
Finished: SUCCESS
```

End of Backup Demo