

Arvados - Idea #8784

[Workbench] Use keep-web to generate directory listings

03/23/2016 06:21 PM - Tom Clegg

Status:	Resolved	Start date:	03/23/2016
Priority:	Normal	Due date:	
Assigned To:	Tom Clegg		
Category:	Keep		
Target version:	2017-07-05 sprint		
Description			
Human- and machine-readable directory listings would allow browsing collections without involving Workbench.			
TBD: Bookmarks to non-public collections in keep-web come with a bit of a snag, which auto-indexing could turn into a bigger practical problem. If you click a bookmark/permalink instead of going through Workbench's links/redirector, and you don't already have an active cookie from <i>previously</i> clicking the Workbench links, there's no option to log in -- you just get an error. You have to recognize the error yourself, find the collection in Workbench, and follow a link from there.			
(Features copied from #5824's "nice to have" section)			
<ul style="list-style-type: none">• Very basic formatting (bootstrap would be nice but can wait)• OK if we show a directory listing only at the top level, similar to the way Workbench does it.• Explicitly no index.html in place of directory listings.<ul style="list-style-type: none">◦ This means "wget -r" can always be expected to work, but it also ensures this can't be used to host entire web sites along the lines of github pages. If we add this capability in the future we could add a -render-index flag. This would make more sense if it came with a bunch of other features anyway, like dynamically mapping vhosts to collections.			
A complementary piece to this is having Workbench generate sharing links which point directly to keep-web rather than redirecting through Workbench.			
This is in support of a request to be able to share data without exposing the Workbench host. https://support.curoverse.com/rt/Ticket/Display.html?id=364			
Subtasks:			
Task # 11887: Review 8784-dir-listings			Resolved
Task # 11872: Workbench redirect			Resolved
Task # 11834: Review 8784-dir-listings			Resolved
Related issues:			
Related to Arvados - Idea #5824: [Workbench] [Keep] collection browse/downloa...			Resolved 05/21/2015
Related to Arvados - Idea #11167: [Workbench] Remove arv-get file download fa...			Resolved 07/24/2017

Associated revisions

Revision 04efddf6 - 06/20/2017 01:45 PM - Tom Clegg

Merge branch '8784-dir-listings'

refs #8784

Arvados-DCO-1.1-Signed-off-by: Tom Clegg <tom@curoverse.com>

Revision a249cd98 - 06/20/2017 02:45 PM - Tom Clegg

8784: Update arvbox to go1.8.

refs #8784

Arvados-DCO-1.1-Signed-off-by: Tom Clegg <tom@curoverse.com>

Revision 0eb72b52 - 07/05/2017 03:39 PM - Tom Clegg

Merge branch '8784-dir-listings'

refs #8784

Arvados-DCO-1.1-Signed-off-by: Tom Clegg <tom@curoverse.com>

Revision 48e8a64b - 04/30/2018 02:03 PM - Tom Clegg

Fix #ZgotmplZ in dir listing page when filename contains a colon.

refs #8784

Arvados-DCO-1.1-Signed-off-by: Tom Clegg <tclegg@veritasgenetics.com>

Revision a6958107 - 04/30/2018 09:02 PM - Tom Clegg

Fix #ZgotmplZ in dir listing page when filename contains a colon.

refs #8784

Arvados-DCO-1.1-Signed-off-by: Tom Clegg <tclegg@veritasgenetics.com>

Revision 643189a6 - 05/01/2018 01:27 PM - Tom Clegg

Fix #ZgotmplZ in dir listing page when filename contains a colon.

refs #8784

Arvados-DCO-1.1-Signed-off-by: Tom Clegg <tclegg@veritasgenetics.com>

History

#1 - 04/05/2016 04:49 PM - Brett Smith

- Target version set to Arvados Future Sprints

#2 - 05/18/2017 03:39 PM - Peter Amstutz

Related to this, workbench sharing links should link directly to the keep-web index page.

#3 - 05/23/2017 07:05 PM - Tom Morris

- Target version changed from Arvados Future Sprints to 2017-06-21 sprint

#4 - 05/30/2017 07:05 PM - Tom Morris

- Description updated

- Story points set to 2.0

#5 - 06/07/2017 06:26 PM - Tom Morris

- Tracker changed from Bug to Idea

#6 - 06/07/2017 06:59 PM - Tom Clegg

- Assigned To set to Tom Clegg

#7 - 06/07/2017 08:28 PM - Tom Clegg

- Status changed from New to In Progress

#8 - 06/13/2017 12:59 PM - Tom Clegg

- File index page 0e3369b.png added

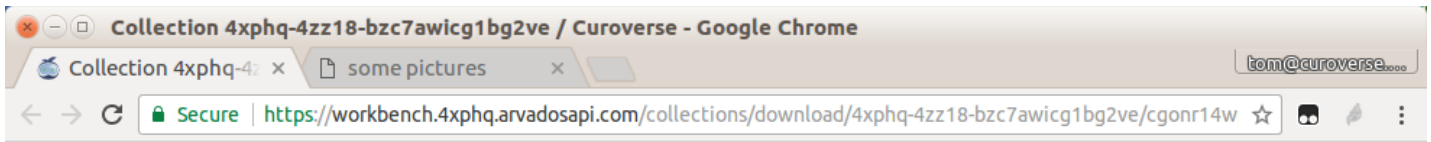
- File index page 2dc0e00.png added

#9 - 06/13/2017 01:00 PM - Tom Clegg

- File index.page.2dc0e00.png added

- File index.page.0e3369b.png added

Existing Workbench index page



Collection 4xphq-4zz18-bzc7awicg1bg2ve

This collection of data files is being shared with you through Arvados. You can download individual files listed below. To download the entire collection with wget, try:

```
$ wget --mirror --no-parent --no-host --cut-dirs=3 https://workbench.4xphq.arvadosapi.com/collections/download/4xphq-4zz18-bzc7awicg1bg2ve/cgonr14w
```

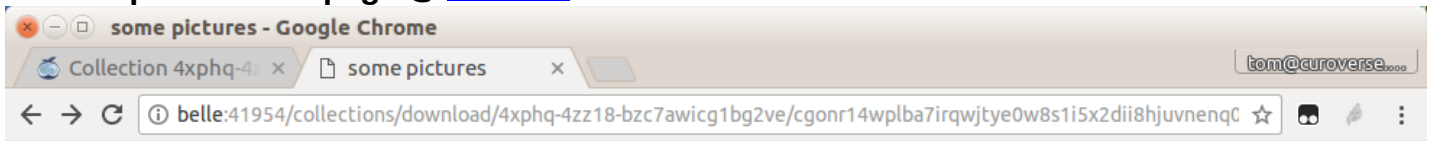
File Listing

- zerp
 - blerp
 - [foo.txt](#)
- [DSC_0108.JPG](#)
- [IMG_20120909_171445.jpg](#)

About Arvados

Arvados is a free and open source software bioinformatics platform. To learn more, visit [arvados.org](#). Arvados is not responsible for the files listed on this page.

new keep-web index page @ [2dc0e00](#)



some pictures

This collection of data files is being shared with you through Arvados. You can download individual files listed below. To download the entire collection with wget, try:

```
$ wget --mirror --no-parent --no-host --cut-dirs=4 https://belle:41954/collections/download/4xphq-4zz18-bzc7awicg1bg2ve/cgonr14wplba7irqwjtje0w8s1i5x2dii8hjuvnenqC
```

File Listing

- [DSC_0108.JPG](#)
- [IMG_20120909_171445.jpg](#)
- [zerp/berp/foo.txt](#)

About Arvados: Arvados is a free and open source software bioinformatics platform. To learn more, visit [arvados.org](#). Arvados is not responsible for the files listed on this page.

TODO:

- use new SDK to load into collection record and cache that, instead of loading/caching a map[string]interface and copying to a Collection in each request

TBD:

- (separate branch) redirect Workbench index page URL to keep-web, and use keep-web directly when generating new links
- monospace font change a good idea?
- dropping the hierarchy a good idea? if so, maybe only the name should be hotlinked, not the whole path?
- rename (*Collection)FileSystem() to ROFileSystem() or HTTPFileSystem() or Snapshot()?

#10 - 06/13/2017 01:01 PM - Tom Clegg

- File deleted (index page 0e3369b.png)

#11 - 06/13/2017 01:01 PM - Tom Clegg

- File deleted (index page 2dc0e00.png)

#12 - 06/13/2017 02:00 PM - Tom Clegg

- File index.page.a42cb73.png added

at [a42cb73](#), with file sizes



some pictures

This collection of data files is being shared with you through Arvados. You can download individual files listed below. To download the entire collection with wget, try:

```
$ wget --mirror --no-parent --no-host --cut-dirs=4 https://belle:41954/collections/download/4xphq-4zz18-bzc7awicg1bg2ve/cgonr14wplba7irqwjtye0w8s1i5x2dii8hjuvnenqC
```

File Listing

1624300	DSC_0108.JPG
946961	IMG_20120909_171445.jpg
4	zerp/blerp/foo.txt

About Arvados: Arvados is a free and open source software bioinformatics platform. To learn more, visit arvados.org. Arvados is not responsible for the files listed on this page.

#13 - 06/13/2017 02:53 PM - Tom Clegg

8784-dir-listings @ [abf007273ba68c2eb541763e40b19d1703132685](#)

#14 - 06/14/2017 03:08 PM - Tom Clegg

8784-dir-listings @ [510a92b885ff547dd7eecb34093f27a7245f021f](#)

- use new SDK to load into collection record and cache that, instead of loading/caching a map[string]interface and copying to a Collection in each request

#15 - 06/14/2017 07:32 PM - Lucas Di Pentima

Sorry for the delay, here are some questions:

- File sdk/go/arvados/collection_fs.go
 - Line 60, 68: Shouldn't ret be checked if it's nil before returning? As I read on File's Readdir doc, if returning an empty slice, it should return a non-nil error.
 - Line 189: I suppose that empty conditional should be removed, right?
- File services/keep-web/handler.go
 - Line 168: Shouldn't that comment line be updated to something like "c=ID[/PATH...]" to stay in sync with the code?
 - Lines 323-335: It's a little difficult (at least for me) to follow that conditionals chain, could you add some comments? I'm specially interested on L329, but in general I feel that for an Open() operation, those are lots of checks, it would be nice to have comments.
 - Line 402: Couldn't collection be retrieved from fs doing type assertion? (to avoid passing it as an argument)
- Don't understand why, but running local tests failed with this:

```
# git.curoverse.com/arvados.git/services/keep-web
tmp/GOPATH/src/git.curoverse.com/arvados.git/services/keep-web/handler.go:451: undefined: sort.Slice
```

- If a collection is empty, is it possible to show some legend stating that fact instead of using the title "File Listing"?

#16 - 06/15/2017 09:26 PM - Tom Clegg

Lucas Di Pentima wrote:

Sorry for the delay, here are some questions:

- File `sdk/go/arvados/collection_fs.go`
 - Line 60, 68: Shouldn't `ret` be checked if it's `nil` before returning? As I read on File's `Readdir` doc, if returning an empty slice, it should return a non-`nil` error.

Yes, updated. If `count>0` then an empty slice has to come with an error.

- Line 189: I suppose that empty conditional should be removed, right?

Whoops. That can't happen: we would have returned it as a file earlier. Removed.

- File `services/keep-web/handler.go`
 - Line 168: Shouldn't that comment line be updated to something like `"/c=ID[/PATH...]"` to stay in sync with the code?

Updated.

- Lines 323-335: It's a little difficult (at least for me) to follow that conditionals chain, could you add some comments? I'm specially interested on L329, but in general I feel that for an `Open()` operation, those are lots of checks, it would be nice to have comments.

Added comments to the error cases.

- Line 402: Couldn't collection be retrieved from `fs` doing type assertion? (to avoid passing it as an argument)

I suppose... turns out the only thing we need from the collection is its name anyway, so I changed that arg to "collectionName string".

- Don't understand why, but running local tests failed with this:
[...]

Ah, it seems that was introduced in Go 1.8 -- I *thought* sorting seemed easier than last time! Updated `run-tests.sh` and packaging scripts accordingly.

- If a collection is empty, is it possible to show some legend stating that fact instead of using the title "File Listing"?

Done.

8784-dir-listings @ [f77d63e6cfaf7278c1cb0fb05e5a4e3f45320e3a](https://ci.curoverse.com/job/8784-dir-listings/177d63e6cfaf7278c1cb0fb05e5a4e3f45320e3a)

#17 - 06/16/2017 04:25 PM - Lucas Di Pentima

LGTM!

#18 - 06/21/2017 03:38 PM - Tom Clegg

8784-dir-listings @ [6c51f11ab5affb4023762227ffb53a5be11a1003](https://ci.curoverse.com/job/8784-dir-listings/6c51f11ab5affb4023762227ffb53a5be11a1003)

- Workbench redirects to keep-web (if available) instead of serving its own directory listing page.
- passed <https://ci.curoverse.com/job/developer-run-tests/342/>

#19 - 06/21/2017 06:36 PM - Tom Clegg

- Subject changed from *[Keep-web] Generate directory listings* to *[Workbench] Use keep-web to generate directory listings*

- Story points changed from 2.0 to 0.5

#20 - 06/21/2017 06:36 PM - Tom Clegg

- Target version changed from 2017-06-21 sprint to 2017-07-05 sprint

#21 - 06/27/2017 08:17 PM - Tom Clegg

8784-dir-listings @ [f723f5f08e286df4c4c58a27a9db574ca604a6f](https://ci.curoverse.com/job/8784-dir-listings/f723f5f08e286df4c4c58a27a9db574ca604a6f) -- passed <https://ci.curoverse.com/job/developer-run-tests/356/>

#22 - 06/29/2017 04:00 PM - Peter Amstutz

What's the rationale for using query_token instead of path_token?

#23 - 06/30/2017 07:41 PM - Tom Clegg

Peter Amstutz wrote:

What's the rationale for using query_token instead of path_token?

None -- changed to path_token.

Also added a fix to make the "upload / network error" integration test pass with the latest Firefox.

8784-dir-listings @ [e59c1d365d9b6e1eff9b5cb030a8b1a3aaf14353](#)

#24 - 07/03/2017 02:07 PM - Peter Amstutz

LGTM @ [e59c1d365d9b6e1eff9b5cb030a8b1a3aaf14353](#)

#25 - 07/05/2017 07:03 PM - Tom Clegg

- Status changed from In Progress to Resolved

Files

index.page.2dc0e00.png	66.6 KB	06/13/2017	Tom Clegg
index.page.0e3369b.png	79.8 KB	06/13/2017	Tom Clegg
index.page.a42cb73.png	67.3 KB	06/13/2017	Tom Clegg