December 2, 2022

The purpose of these procedures is to document the process to work in the docker environment for Galaxy, since I think that it is unstable and tricky and want everyone in our Group to know how to work with it.
This document will be included in the galaxy-upgrade git repository, in the hutlab-tools directory:
https://github.com/biobakery/galaxy-upgrade/tree/main/hutlab_tools

Please note that Galaxy provides excellent, extensive documentation on how to develop new tools, for example:

- https://galaxyproject.org/admin/tools/add-tool-tutorial/
- https://docs.galaxyproject.org/en/master/dev/index.html
- https://docs.galaxyproject.org/en/master/dev/schema.html

The procedures below are *focused on the particular Test Galaxy Docker environment we use in the Huttenhower Lab*.

I would like to emphasize that there could be shortcuts, improvements, etc. but my focus is to share a solid procedure that works for me and that **ensures the integrity** of the changes in the volatile galaxy / docker environment we have.

Prepared by George Weingart, Dec. 2, 2022

Important notes:

1. Our code is located in **https://github.com/biobakery/galaxy-upgrade.git**
2. The Galaxy test site is located in http://jdrf-mibc.org/galaxy-test/

To demonstrate the procedures, I will use the following simple exercise:

We will add "sam" to the allowable input types for humann.

Looking at "humann.xml" we see the following code:

```
<inputs>
      <param type="data" name="input" format="fasta, fastq" />
</inputs>
```

We want to modify the xml to allow for sam   as follows:

```
<inputs>
      <param type="data" name="input" format="fasta, fastq,  sam" />
</inputs>
```

Procedure:

Login:

ssh <Your.ID>@login.rc.fas.harvard.edu

Login to jdrf2:

ssh jdrf2

Login to the container:

Select the name of the container:
sudo docker ps -a

Select the right container, in our case galaxy_updated_container_v1

Login to the container:

sudo docker exec -it galaxy_updated_container_v1 bash

Go to the tools directory:

cd tools

IMPORTANT:

Our code is located in the following library:

galaxy-upgrade

We will verify we are current with the code in git:

```
(base) root@5cc594913c8f:/galaxy-central/tools/galaxy-upgrade# pwd
/galaxy-central/tools/galaxy-upgrade
(base) root@5cc594913c8f:/galaxy-central/tools/galaxy-upgrade# git pull
Already up to date.
(base) root@5cc594913c8f:/galaxy-central/tools/galaxy-upgrade#
```

Perform the update

Go to the correct library, in this case humann

cd humann

vim humann.xml

And modify the text we want to change:

<inputs>

```
        <param type="data" name="input" format="fasta, fastq, sam" />
</inputs>
```

Save the file.

We will safeguard the file now by pushing it up to git

(base) root@5cc594913c8f:/galaxy-central/tools/galaxy-upgrade/humann# **git status**
On branch main
Your branch is up to date with 'origin/main'.

Changes not staged for commit:
  (use "git add <file>..." to update what will be committed)
  (use "git checkout -- <file>..." to discard changes in working directory)

        modified: humann.xml

Untracked files:
  (use "git add <file>..." to include in what will be committed)

        demo.fasta
        test1.sh
        ../melonnpan/
        ../panphlan/panphlan_backup.xml
        ../panphlan/panphlan_wrapper_Backup.sh

no changes added to commit (use "git add" and/or "git commit -a")
(base) root@5cc594913c8f:/galaxy-central/tools/galaxy-upgrade/humann# **git add humann.xml**
(base) root@5cc594913c8f:/galaxy-central/tools/galaxy-upgrade/humann# **git commit -m "Updated humann.xml by adding file type sam to input"**
[main 2498adc] Updated humann.xml by adding file type sam to input
 1 file changed, 1 insertion(+), 1 deletion(-)
(base) root@5cc594913c8f:/galaxy-central/tools/galaxy-upgrade/humann# **git push -u origin main**
Username for 'https://github.com': George-Weingart
Password for 'https://George-Weingart@github.com':
Counting objects: 4, done.
Delta compression using up to 48 threads.
Compressing objects: 100% (4/4), done.
Writing objects: 100% (4/4), 372 bytes | 372.00 KiB/s, done.
Total 4 (delta 3), reused 0 (delta 0)

remote: Resolving deltas: 100% (3/3), completed with 3 local objects.
To https://github.com/biobakery/galaxy-upgrade.git
   4b8cfca..2498adc main -> main
Branch 'main' set up to track remote branch 'main' from 'origin'.
(base) root@5cc594913c8f:/galaxy-central/tools/galaxy-upgrade/humann#

Sometimes, it is enough to recycle Galaxy in the container, but more often than not, I have had to recycle the container, refresh the apps in tool_conf.xml and recycle galaxy.

We now copy the xml to the executable library /galaxy-central/tools/humann:

(base) root@5cc594913c8f:/galaxy-central/tools/galaxy-upgrade/humann# **pwd**
/galaxy-central/tools/galaxy-upgrade/humann
(base) root@5cc594913c8f:/galaxy-central/tools/galaxy-upgrade/humann# **cp humann.xml ../../humann**
(base) root@5cc594913c8f:/galaxy-central/tools/galaxy-upgrade/humann#

Recycle the container and Galaxy

Exit the container:

(base) root@5cc594913c8f:/galaxy-central/tools/humann# **exit**
exit

Verify the container:

sudo docker ps -a

Restart it:

sudo docker restart galaxy_updated_container_v1

Login to the container:

sudo docker exec -it galaxy_updated_container_v1 bash

IMPORTANT

We need to update the applications - add **our applications** to **config/tool_conf.xml:**

List of our applications is located in

/galaxy-central/tools/galaxy-upgrade/hutlab_tools/tool_conf_additions.xml

We copy the contents of that file:

```
<section name="LEfSe" id="lefse">
  <tool file="galaxy_lefse/format_input.xml" />
  <tool file="galaxy_lefse/run_lefse.xml" />
  <tool file="galaxy_lefse/plot_res.xml" />
  <tool file="galaxy_lefse/plot_cladogram.xml" />
  <tool file="galaxy_lefse/plot_single_feature.xml" />
  <tool file="galaxy_lefse/plot_features.xml" />
 </section>
 <section name="GraPhlAn" id="graphlan">
  <tool file="galaxy_graphlan/graphlan_annotate.xml" />
  <tool file="galaxy_graphlan/graphlan_ring_annotate.xml" />
  <tool file="galaxy_graphlan/graphlan.xml" />
 </section>
 <section id="MetaPhLan" name="MetaPhLan">
    <tool file="metaphlan/metaphlan.xml" />
 </section>
 <section id="MaAsLin" name="MaAsLin">
    <tool file="maaslin/maaslin.xml" />
 </section>
 <section id="HUMAnN" name="HUMAnN">
   <tool file="humann/humann.xml" />
 </section>
 <section id="PanPhlAn" name="PanPhlAn">
    <tool file="panphlan/panphlan.xml" />
 </section>
 <section id="SparseDOSSA" name="SparseDOSSA">
    <tool file="sparsedossa/sparsedossa.xml" />
 </section>
  <section id="MMUPHin" name="MMUPHin">
```

```
      <tool file="mmuphin/mmuphin_preprocess_metadata.xml" />
      <tool file="mmuphin/adjust_batch.xml" />
      <tool file="mmuphin/lm_meta.xml" />
 </section>
 <section id="HAllA" name="HAllA">
      <tool file="galaxy_halla/halla.xml" />
 </section>
 <section id="WAAFLE" name="WAAFLE">
      <tool file="galaxy_waafle/waafle1.xml" />
      <tool file="galaxy_waafle/waafle2.xml" />
      <tool file="galaxy_waafle/waafle3.xml" />
 </section>
```

 in an appropriate location in **/galaxy-central/config/tool_conf.xml**

(base) root@5cc594913c8f:/galaxy-central/tools/galaxy-upgrade/hutlab_tools# **cd ../../../config**

(base) root@5cc594913c8f:/galaxy-central/config# **vim tools_conf.xml**

The changed part of this file, tool.conf.xml, will look as follows:

WARNING

Make sure the syntax of that file does not get corrupted, because if it does, galaxy restart fails with a "spawn error".
Inversely, if you get a "spawn error" when you restart galaxy, the error is probably you messed up the xml syntax of the file (Missing </section?? Or Duplicate </section ?? etc.)

```
  <tool file="data_source/zebrafishmine.xml" />
  <tool file="data_source/eupathdb.xml" />
  <tool file="data_source/hbvar.xml" />
 </section>
 <section id="send" name="Send Data">
  <tool file="cloud/send.xml" />
 </section>
<section name="LEfSe" id="lefse">
  <tool file="galaxy_lefse/format_input.xml" />
```

```xml
    <tool file="galaxy_lefse/run_lefse.xml" />
    <tool file="galaxy_lefse/plot_res.xml" />
    <tool file="galaxy_lefse/plot_cladogram.xml" />
    <tool file="galaxy_lefse/plot_single_feature.xml" />
    <tool file="galaxy_lefse/plot_features.xml" />
  </section>
  <section name="GraPhlAn" id="graphlan">
    <tool file="galaxy_graphlan/graphlan_annotate.xml" />
    <tool file="galaxy_graphlan/graphlan_ring_annotate.xml" />
    <tool file="galaxy_graphlan/graphlan.xml" />
  </section>
  <section id="MetaPhLan" name="MetaPhLan">
      <tool file="metaphlan/metaphlan.xml" />
  </section>
  <section id="MaAsLin" name="MaAsLin">
      <tool file="maaslin/maaslin.xml" />
  </section>
  <section id="HUMAnN" name="HUMAnN">
    <tool file="humann/humann.xml" />
  </section>
  <section id="PanPhlAn" name="PanPhlAn">
      <tool file="panphlan/panphlan.xml" />
  </section>
  <section id="SparseDOSSA" name="SparseDOSSA">
      <tool file="sparsedossa/sparsedossa.xml" />
  </section>
  <section id="MMUPHin" name="MMUPHin">
       <tool file="mmuphin/mmuphin_preprocess_metadata.xml" />
      <tool file="mmuphin/adjust_batch.xml" />
      <tool file="mmuphin/lm_meta.xml" />
  </section>
  <section id="HAllA" name="HAllA">
       <tool file="galaxy_halla/halla.xml" />
  </section>
  <section id="WAAFLE" name="WAAFLE">
      <tool file="galaxy_waafle/waafle1.xml" />
      <tool file="galaxy_waafle/waafle2.xml" />
      <tool file="galaxy_waafle/waafle3.xml" />
  </section>

  <section id="collection_operations" name="Collection Operations">
  <section id="collection_operations" name="Collection Operations">
    <tool file="${model_tools_path}/unzip_collection.xml" />
    <tool file="${model_tools_path}/zip_collection.xml" />
    <tool file="${model_tools_path}/filter_failed_collection.xml" />
    <tool file="${model_tools_path}/filter_empty_collection.xml" />
```

```xml
    <tool file="${model_tools_path}/flatten_collection.xml" />
    <tool file="${model_tools_path}/merge_collection.xml" />
    <tool file="${model_tools_path}/relabel_from_file.xml" />
    <tool file="${model_tools_path}/filter_from_file.xml" />
    <tool file="${model_tools_path}/sort_collection_list.xml" />
    <tool file="${model_tools_path}/tag_collection_from_file.xml" />
    <tool file="${model_tools_path}/apply_rules.xml" />
    <tool file="${model_tools_path}/build_list.xml" />
    <tool file="${model_tools_path}/extract_dataset.xml" />
  </section>
  <section id="expression_tools" name="Expression Tools">
 </section>
```

Recycle Galaxy in the container:

(base) root@5cc594913c8f:/galaxy-central/config# **supervisorctl restart galaxy:**
galaxy:galaxy_nodejs_proxy: stopped
galaxy:handler1: stopped
galaxy:handler0: stopped
galaxy:galaxy_web: stopped
galaxy:galaxy_nodejs_proxy: started
galaxy:galaxy_web: started
galaxy:handler0: started
galaxy:handler1: started
(base) root@5cc594913c8f:/galaxy-central/config#

Test the change:

The expectation is that when we bring up Galaxy, it provides the option to process a sam dataset (In addition to fasta and fastq that were there before) and indeed that is the case - change successful !

Logs:

In case you'd like to see the logs of galaxy, they are located in **/home/galaxy/logs**

The most important is uwsgi.log

```
76.126.124.30 - - [18/May/2022:19:07:35 +0000] "POST /galaxy-test/api/tools/HUMAnN/build HTTP/1.1" 200 - "http://jdrf-mibc.org/galaxy-test" "Mozilla/5.0 (Macintosh; Intel Mac O
S X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/101.0.4951.64 Safari/537.36"
[pid: 3780|app: 0|req: 13/22] 172.17.0.3 () {58 vars in 1119 bytes} [Wed May 18 19:07:35 2022] POST /api/tools/HUMAnN/build => generated 6370 bytes in 113 msecs (HTTP/1.1 200)
3 headers in 139 bytes (1 switches on core 0)
galaxy.tools INFO 2022-05-18 19:07:36,858 [p:3783,w:2,m:0] [uWSGIWorker2Core2] Validated and populated state for tool request (77.379 ms)
galaxy.tools.actions INFO 2022-05-18 19:07:36,957 [p:3783,w:2,m:0] [uWSGIWorker2Core2] Handled output named output1 for tool HUMAnN (60.836 ms)
galaxy.tools.actions INFO 2022-05-18 19:07:36,984 [p:3783,w:2,m:0] [uWSGIWorker2Core2] Handled output named output2 for tool HUMAnN (26.831 ms)
galaxy.tools.actions INFO 2022-05-18 19:07:37,006 [p:3783,w:2,m:0] [uWSGIWorker2Core2] Handled output named output3 for tool HUMAnN (22.508 ms)
galaxy.tools.actions INFO 2022-05-18 19:07:37,026 [p:3783,w:2,m:0] [uWSGIWorker2Core2] Added output datasets to history (19.300 ms)
galaxy.tools.actions INFO 2022-05-18 19:07:37,031 [p:3783,w:2,m:0] [uWSGIWorker2Core2] Setup for job Job[unflushed,tool_id=HUMAnN] complete, ready to be enqueued (4.562 ms)
galaxy.jobs.manager DEBUG 2022-05-18 19:07:37,031 [p:3783,w:2,m:0] [uWSGIWorker2Core2] (Job[unflushed,tool_id=HUMAnN]) Configured job handler for tool 'HUMAnN' is: handlers
galaxy.web_stack.handlers DEBUG 2022-05-18 19:07:37,031 [p:3783,w:2,m:0] [uWSGIWorker2Core2] (Job[unflushed,tool_id=HUMAnN]) Selected handler 'handler1' by random choice from h
andler tag 'handlers'
galaxy.web_stack.handlers INFO 2022-05-18 19:07:37,125 [p:3783,w:2,m:0] [uWSGIWorker2Core2] Flushed transaction for Job[id=2093,tool_id=HUMAnN] (93.204 ms)
galaxy.web_stack.handlers INFO 2022-05-18 19:07:37,125 [p:3783,w:2,m:0] [uWSGIWorker2Core2] (Job[id=2093,tool_id=HUMAnN]) Handler 'handler1' assigned using 'db-preassign' assig
nment method
galaxy.tools.execute DEBUG 2022-05-18 19:07:37,126 [p:3783,w:2,m:0] [uWSGIWorker2Core2] Tool HUMAnN created job 2093 (246.245 ms)
galaxy.tools.execute DEBUG 2022-05-18 19:07:37,141 [p:3783,w:2,m:0] [uWSGIWorker2Core2] Executed 1 job(s) for tool HUMAnN request (281.985 ms)
76.126.124.30 - - [18/May/2022:19:07:36 +0000] "POST /galaxy-test/api/tools HTTP/1.1" 200 - "http://jdrf-mibc.org/galaxy-test" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7)
AppleWebKit/537.36 (KHTML, like Gecko) Chrome/101.0.4951.64 Safari/537.36"
[pid: 3783|app: 0|req: 10/23] 172.17.0.3 () {58 vars in 1093 bytes} [Wed May 18 19:07:36 2022] POST /api/tools => generated 2955 bytes in 426 msecs (HTTP/1.1 200) 3 headers in
139 bytes (1 switches on core 2)
76.126.124.30 - - [18/May/2022:19:07:37 +0000] "GET /galaxy-test/api/webhooks HTTP/1.1" 200 - "http://jdrf-mibc.org/galaxy-test" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7
) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/101.0.4951.64 Safari/537.36"
[pid: 3783|app: 0|req: 12/24] 172.17.0.3 () {52 vars in 981 bytes} [Wed May 18 19:07:37 2022] GET /api/webhooks => generated 2 bytes in 14 msecs (HTTP/1.1 200) 3 headers in 139
 bytes (1 switches on core 3)
76.126.124.30 - - [18/May/2022:19:07:37 +0000] "GET /galaxy-test/api/histories/591b5307871a50b3/contents?order=hid&v=dev&q=update_time-ge&q=deleted&q=purged&qv=2022-05-18T19%3A
02%3A01.000Z&qv=False&qv=False HTTP/1.1" 200 - "http://jdrf-mibc.org/galaxy-test" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome
/101.0.4951.64 Safari/537.36"
[pid: 3783|app: 0|req: 12/25] 172.17.0.3 () {52 vars in 1238 bytes} [Wed May 18 19:07:37 2022] GET /api/histories/591b5307871a50b3/contents?order=hid&v=dev&q=update_time-ge&q=d
eleted&q=purged&qv=2022-05-18T19%3A02%3A01.000Z&qv=False&qv=False => generated 1979 bytes in 57 msecs (HTTP/1.1 200) 3 headers in 139 bytes (1 switches on core 1)
76.126.124.30 - - [18/May/2022:19:07:45 +0000] "GET /galaxy-test/api/histories/591b5307871a50b3/contents?order=hid&v=dev&q=update_time-ge&q=deleted&q=purged&qv=2022-05-18T19%3A
07%3A37.000Z&qv=False&qv=False HTTP/1.1" 200 - "http://jdrf-mibc.org/galaxy-test" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome
/101.0.4951.64 Safari/537.36"
[pid: 3780|app: 0|req: 14/26] 172.17.0.3 () {52 vars in 1238 bytes} [Wed May 18 19:07:45 2022] GET /api/histories/591b5307871a50b3/contents?order=hid&v=dev&q=update_time-ge&q=d
eleted&q=purged&qv=2022-05-18T19%3A07%3A37.000Z&qv=False&qv=False => generated 1991 bytes in 43 msecs (HTTP/1.1 200) 3 headers in 139 bytes (1 switches on core 1)
76.126.124.30 - - [18/May/2022:19:07:50 +0000] "GET /galaxy-test/api/histories/591b5307871a50b3/contents?order=hid&v=dev&q=update_time-ge&q=deleted&q=purged&qv=2022-05-18T19%3A
07%3A45.000Z&qv=False&qv=False HTTP/1.1" 200 - "http://jdrf-mibc.org/galaxy-test" "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome
/101.0.4951.64 Safari/537.36"
[pid: 3780|app: 0|req: 15/27] 172.17.0.3 () {52 vars in 1238 bytes} [Wed May 18 19:07:50 2022] GET /api/histories/591b5307871a50b3/contents?order=hid&v=dev&q=update_time-ge&q=d
eleted&q=purged&qv=2022-05-18T19%3A07%3A45.000Z&qv=False&qv=False => generated 2 bytes in 37 msecs (HTTP/1.1 200) 3 headers in 139 bytes (1 switches on core 3)
```