



Managing  
dependencies



Managing and executing  
analysis workflow



Versioning and  
collaborating on code  
(and some other files)



Connecting code  
and reporting

and...



Isolating and exporting  
environment

Results should be possible to reproduce  
regardless of platform and with minimal effort.



“Docker provides a way to run applications securely  
isolated in a container, packaged with all its  
dependencies and libraries.”

We report extensive metagenomic and metatranscriptomic profiling of the microbial community from a production-scale biogas plant. Given the unprecedented sequencing depth and established bioinformatics, our data will be of great interest to the biogas research community in general and microbiologists working on biogas-producing microbial communities in particular. In a first applied study, our metagenome assembly was used to improve the characterization of a metaproteome generated from biogas plant fermentation samples and to investigate the metabolic activity of the microbial community [17].

## Availability of supporting data

Raw sequencing data are available in the European Nucleotide Archive (ENA) under study accession PRJEB8813 (<http://www.ebi.ac.uk/ena/data/view/PRJEB8813>). The datasets supporting the results of this article are available in *GigaScience's* GigaDB [2].

The complete workflow is organized in a single GNU Makefile and available on GitHub [18]. All data and results can be reproduced by a simple invocation of *make*. To further support reproducibility, we bundled all tools and dependencies into one Docker container available on DockerHub [19]. *docker run* executes the aforementioned Makefile inside the container. Reproduction

```
$uname -a
```

```
Darwin dhcp-140-26.vpn.chalmers.se 15.6.0 Darwin Kernel Version 15.6.0:  
Thu Sep  1 15:01:16 PDT 2016; root:xnu-3248.60.11~2/RELEASE_X86_64  
x86_64
```

```
$docker pull ubuntu:16.04
```

```
16.04: Pulling from library/ubuntu
```

```
22dc81ace0ea: Pull complete
```

```
1a8b3c87dba3: Pull complete
```

```
91390a1c435a: Pull complete
```

```
07844b14977e: Pull complete
```

```
b78396653dae: Pull complete
```

```
Digest:
```

```
sha256:e348fbbbea0e0a0e73ab0370de151e7800684445c509d46195aef73e090a49bd6
```

```
Status: Downloaded newer image for ubuntu:16.04
```

```
$docker run -it ubuntu:16.04
```

```
root@407b0fd13fe5:/# uname -a
```

```
Linux 407b0fd13fe5 4.9.60-linuxkit-aufs #1 SMP Mon Nov 6 16:00:12 UTC  
2017 x86_64 x86_64 x86_64 GNU/Linux
```

## Dockerfile

```
FROM ubuntu:16.04

# Install prerequisites
RUN apt-get update && \
    apt-get install -y --no-install-recommends \
    bzip2 curl ca-certificates

# Install Conda
RUN curl https://repo.continuum.io/miniconda.sh -O && \
    bash miniconda.sh -bf -p /opt/miniconda3/ && \
    rm miniconda.sh

# Add conda to PATH
ENV PATH="/opt/miniconda3/bin:${PATH}"

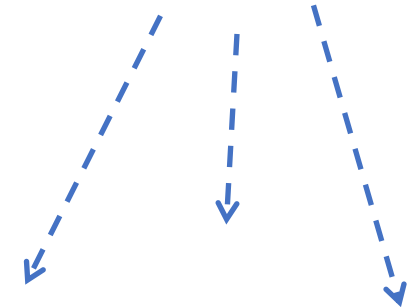
# Install git and nano from conda-forge
RUN conda install -c conda-forge git nano

# Use bash as shell
SHELL ["/bin/bash", "-c"]

# Set workdir
WORKDIR /home
```

Build

Docker image



Docker container

Docker container

Docker container

# Mounting volumes

## Local project directory

```
project
|- doc/
|
|- data/
|   |- raw_external/
|   |- raw_internal/
|   |- meta/
|
|- code/
|- notebooks/
|
|- intermediate/
|- scratch/
|- logs/
|
|- results/
|   |- figures/
|   |- tables/
|   |- reports/
|
|- Snakefile
|- config.yml
|- environment.yml
|- Dockerfile
```

## Docker image file system

```
|- bin/
|- boot/
|- dev/
|- etc/
|- home/
|   |- data/
|       |- raw_external/
|       |- raw_internal/
|       |- meta/
|- lib/
|- lib64/
|- media/
|- opt/
|- proc/
|- root/
|- run/
|- sys/
|- tmp/
|- usr/
|- var/
```

```
$docker run -it -v $PWD/data:/home/data ubuntu:16.04
```

# What can I use Docker for?

As an advanced environment manager.



```
$docker run -it  
-v $PWD:/home  
my_image /bin/bash
```

To package your code with the environment it needs.



```
$docker run  
-v $PWD/data:/home/data  
-v $PWD/results:/home/results  
my_image snakemake report.pdf
```

To package a whole workflow, e.g. to accompany a manuscript.



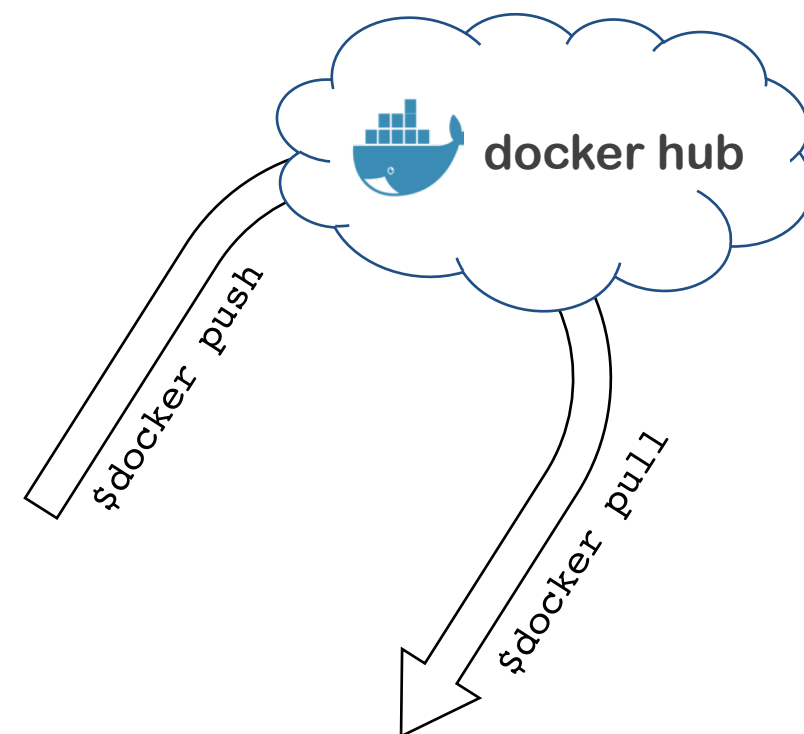
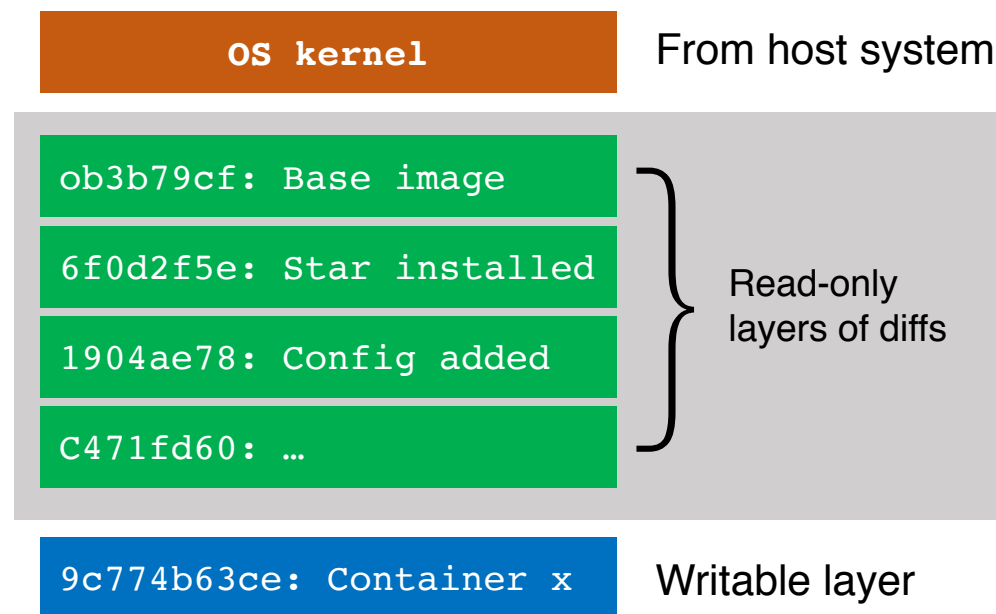
```
$docker run  
-v $PWD/results:/home/results  
my_image snakemake report.pdf
```

## A Docker image

Dockerfile

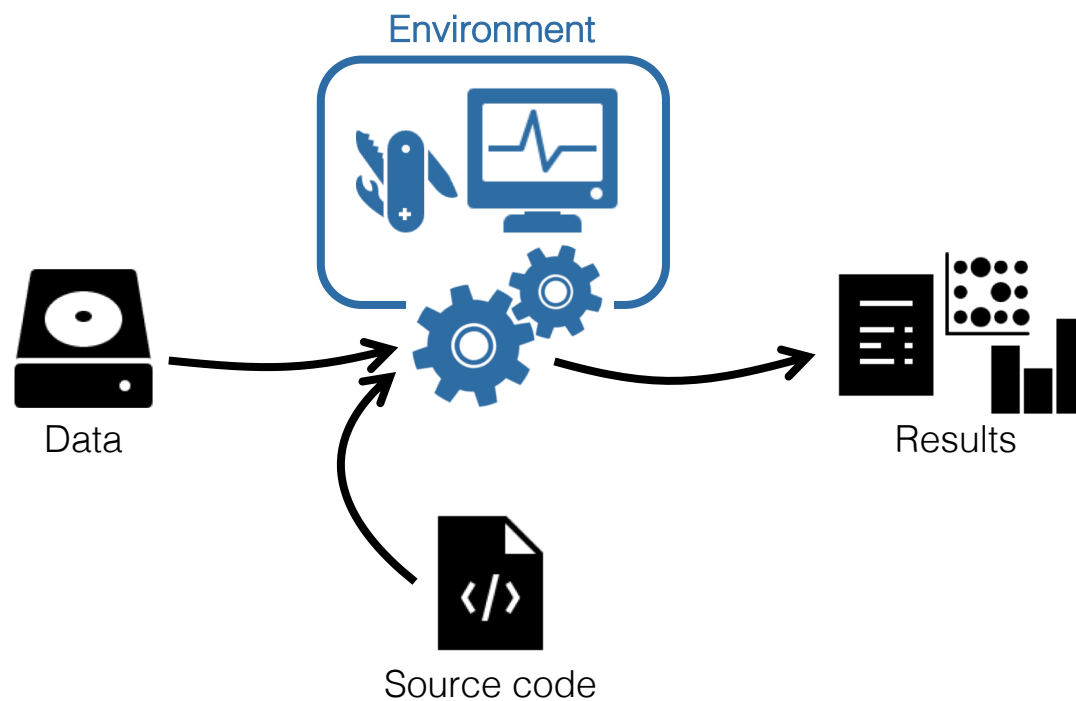
```
FROM debian:latest
RUN conda install star
ADD config.yml
...
```

me\$ docker run image\_id



you\$ docker run image\_id





```
project
|- doc/
|
|- data/
|   |- raw_external/
|   |- raw_internal/
|   |- meta/
|
|- code/
|- notebooks/
|
|- intermediate/
|- scratch/
|- logs/
|
|- results/
|   |- figures/
|   |- tables/
|   |- reports/
|
|- Snakefile
|- config.yml
|- environment.yml
|- Dockerfile
```