Data Repositories for everyday use

Carl Boettiger
Daily workflow

Published data

inputs: CSV → code → outputs: CSV
Daily workflow

Published data

inputs → code → outputs

CSV → CSV
Daily workflow

Different Tools

Similar Tasks

Published data
Published data

Daily workflow

Different Tools

Similar Tasks

- Share data between different users + machines
- Metadata
- Track versions (data + software)
- Find previous work
- Backups & persistent storage
Daily workflow

Published data

Different Tools
Similar Tasks

- Share data between different users + machines
- Metadata
- Track versions (data + software)
- Find previous work
- Backups & persistent storage

- Share data between different users + machines
- Metadata
- Track versions (data + software)
- Find previous work
- Backups & persistent storage
Can’t we all share?
What’s in a name?

vostok.icecore.co2

https://github.com/espm-157/climate-template/releases/download/data/vostok.icecore.co2

https://doi.org/10.3334/CDIAC/ATG.009
What’s in a name?

vostok.icecore.co2

https://github.com/espm-157/climate-template/releases/download/data/vostok.icecore.co2

hash://sha256/9412325831dab2

https://doi.org/10.3334/CDIAC/ATG.009
Location-agnostic names?

hash://sha256/9412325831dab2

- Frictionless identifier
- No authentication, fee, or special software
- Works locally, on cloud, on data archives
- Cryptographically secure
Location-agnostic names?

hash://sha256/9412325831dab2

c contentid::query_sources("hash://sha256/9412325831dab2")
Location-agnostic names?

hash://sha256/9412325831dab2

contentid::query_sources("hash://sha256/9412325831dab2")

```r
#> source
#> 1 /home/rstudio/.local/share/contentid/data/94/12/9412325831dab22aeebdd674b6eb53ba6b7bdd04bb99a4dbb21ddf646287e37
#> 2 https://archive.softwareheritage.org/api/1/content/sha256:9412325831dab22aeebdd674b6eb53ba6b7bdd04bb99a4dbb21ddf646287e37/raw/
#> 3 https://knb.ecoinformatics.org/knb/d1/mn/v2/object/ess-dive-457358fdc81d3a5-20180726T203952542
#> 4 https://github.com/espm-157/climate-template/releases/download/data/vostok.icecore.co2
#> 5 https://minio.thelio.carlboettiger.info/shared-data/vostok.icecore.co2
#> 6 https://cdiac.ess-dive.lbl.gov/ftp/trends/co2/vostok.icecore.co2
```
Daily workflow

Different Tools

Similar Tasks

Published data

EML
PROV
RDF
BAGIT
REST API