
FAIR
vs.
PhyloForge

F.A.I.R'S PRINCIPLES

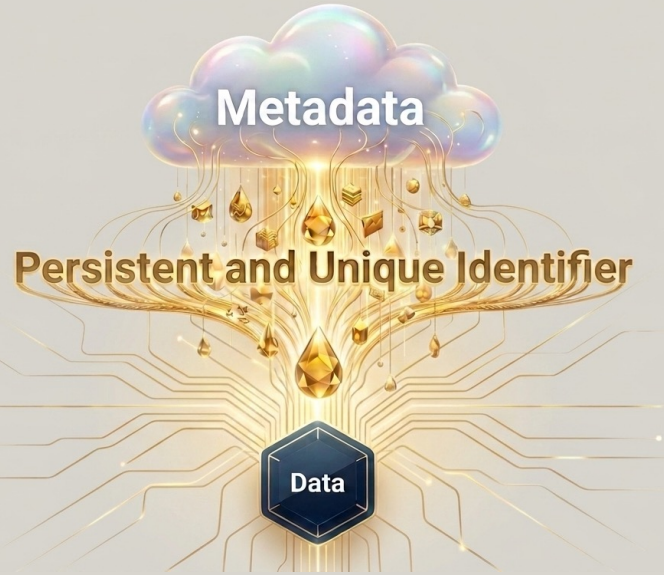
- Findable → *"(Meta)data should be human- and machine-readable"*
- Accessible → *"The user needs to know how the required data can be accessed"*
- Interoperable → *"Data must support integration and interoperability across datasets, applications, and workflows"*
- Reusable → *"Metadata and data should be documented for replication and use in different settings"*

Each and every one of them must be respected

for good science to be made

FINDABLE

- F1: (Meta)data are assigned a globally unique and persistent identifier
- F2: Data are described with rich metadata
- F3: Metadata clearly and explicitly include the identifier of the data they describe
- F4: (Meta)data are registered or indexed in a searchable resource



ACCESSIBLE

- A1: (Meta)data are retrievable by their identifier using a standardized communication protocol
  - A1.1: The protocol is open, free, and universally implementable
  - A1.2: The protocol allows for authentication and authorization procedure where necessary
- A2: (Meta)data are accessible, even when the data is no longer available



INTEROPERABLE

- I1: (Meta)data use a formal, accessible, shared and broadly applicable language for knowledge representation
- I2: (Meta)data use vocabularies that follow FAIR principles
- I3: (Meta)data include qualified references to other (meta)data



REUSABLE

- R1: (Meta)data are richly described with a plurality of accurate and relevant attributes
  - R1.1: (Meta)data are released with a clear and accessible data usage license
  - R1.2: (Meta)data are associated with detailed provenance
  - R1.3: (Meta)data meet domain-relevant community standards



A FEW NOTES ON THE F.A.I.R PRINCIPLES…

- There is a lot of redundancy between them (e.g. about how Metadata need to be rich)
- They are written mostly with regards to how PUBLISHED data should be organised

PHYLOFORGE VS. FINDABILITY

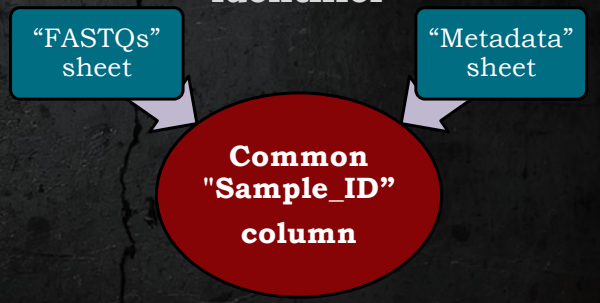
"(Meta)data should be human- and machine-readable"

- Fixed starting point for PhyloForge: an **Excel** file with a **rigid structure**
  - 1 "FASTQs" sheet
  - 1 "Metadata" sheet (*optional* but expected)

  } both included in final **report**
- Common **identifier** = Sample_ID column
  + Error detection mechanisms to ensure the Sample_IDs are coherent between the sheets
- **F2's Rich metadata** = user's responsibility
- Metadata explorable in PhyloForge

- (Meta)data are assigned a globally unique and persistent **identifier**
- Data are described with **rich metadata**

Globally unique and persistent **identifier**

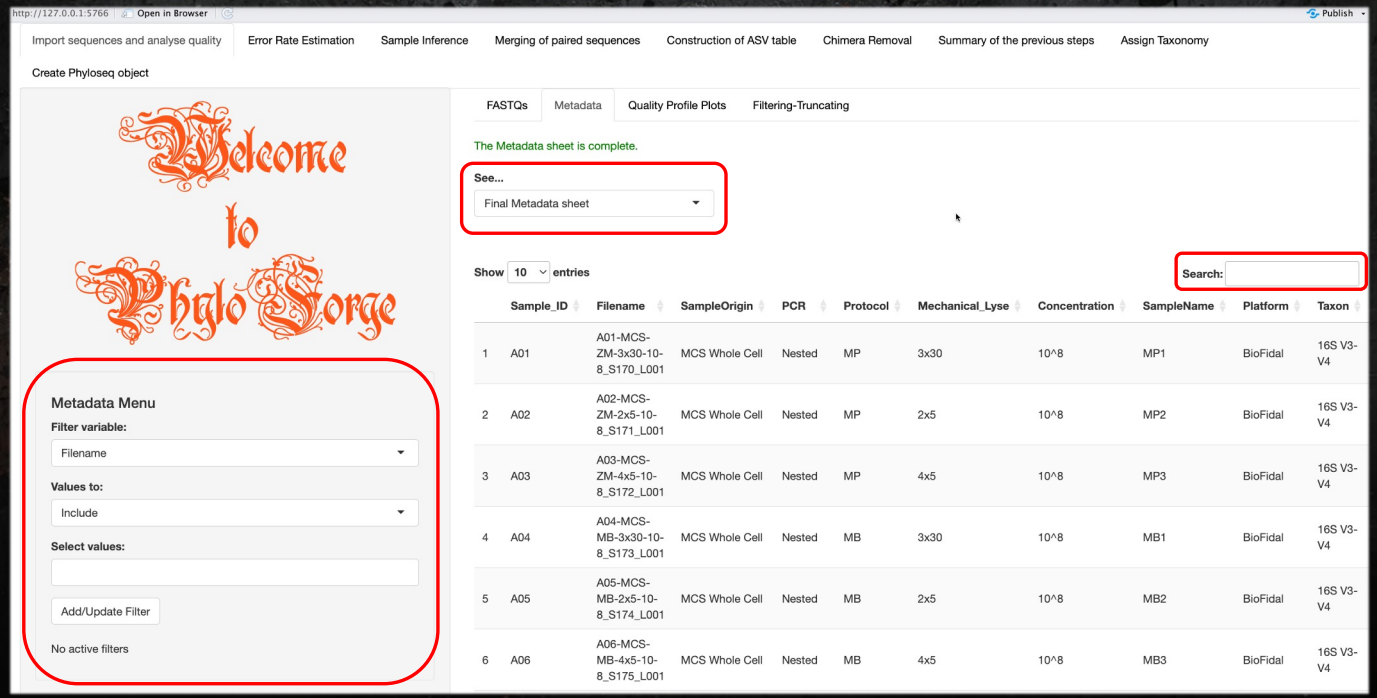


- Metadata clearly and explicitly **include the identifier** of the data they describe
- (Meta)data are registered or indexed in a searchable resource

PHYLOFORGE VS. FINDABILITY

"(Meta)data should be human- and machine-readable"

👍 Metadata explorable in PhyloForge:

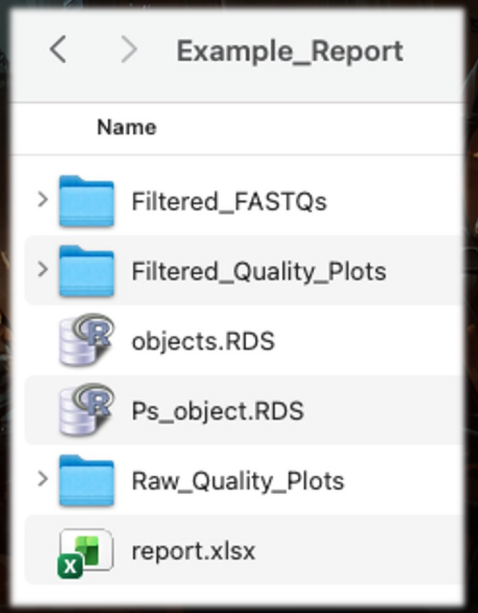


PHYLOFORGE VS. FINDABILITY

"(Meta)data should be human- and machine-readable"

- At the end: Produces a comprehensive and squarely formatted Excel report (+ optional add-ons: plots, intermediate R objects…)
- **Everything** is **in the same place** ➔ the zipped folder can be published "as is" alongside other scientific resources
- And the Excel report (and R object) can be submitted to PhyloForge: (meta)data exploration facilitated here, too

Also contains original data and metadata sheets



PHYLOFORGE VS. ACCESSIBILITY

👍 PhyloForge is a free, open-source application

👍 PhyloForge generates a **report** = ALL the inputs and outputs of the analysis

➔ the protocol and the metadata are **in the same place**

👍 The "FASTQs" sheet lists file **paths**: relative to a specific machine

➔ if the FASTQs themselves are not published, the dataset is unavailable; even so, the Metadata sheet remains complete in the report

➔ Accessibility guidelines are mostly about the way data ought to be **published**, which depends on the **user**, not PhyloForge

⚔ (Meta)data are retrievable by their identifier using a standardized communication protocol 👍

⚔ The protocol is open, free, and universally implementable 👍

⚔ The protocol allows for authentication and authorization procedure, where necessary 👎

➔ Not necessarily 'open' or 'free', but accessibility is documented

⚔ (Meta)data are accessible, even when the data is no longer available 👍

PHYLOFORGE VS. INTEROPERABILITY

"Data must support integration and interoperability across datasets, applications, and workflows"

- The required Excel format is straightforward for machines and humans, but PhyloForge is the **only** tool that expects this EXACT format
- The Metadata sheet is to be filled in by the **user**: PhyloForge cannot do it for you
- PhyloForge produces a **Phyloseq object**, which is processable by the PhyloIgSeq app, or general Phyloseq workflows.

➔ The **core** principle of Interoperability is followed

- (Meta)data use a formal, accessible, shared and broadly applicable language for knowledge representation
  ➔ Data should be readily readable by machines
- (Meta)data use vocabularies that follow FAIR principles
- (Meta)data include qualified references to other (meta)data
  ➔ It should be clear **why** you're referencing something

PHYLOFORGE VS. REUSABLE

"Metadata and data should be documented for replication and use in different settings"

- Again, Metadata completeness and accuracy are the **user's responsibility**, as are **legal** aspects
- The Metadata sheet is an Excel sheet, not a written article: you cannot reference things or write explanations in it, because it would clutter the Excel
- PhyloForge follows the DADA2 standards; the Excel report contains recognisable DADA2 parameters and outputs
- PhyloForge generates an Excel report, containing the protocol used: PhyloForge analyses are **reproducible, i.e. reusable**

- (Meta)data are richly described with a plurality of accurate and relevant attributes
- (Meta)data are released with a clear and accessible data usage license
  ➔ Is about *legal* interoperability
- (Meta)data are associated with detailed provenance
- (Meta)data meet domain-relevant community standards

Overall Results

PhyloForge follows the general F.A.I.R principles.

1) PhyloForge is a FAIR-based system to support a FAIR-based publication approach
2) Most of the guidelines are about the presence and completeness of Metadata: but only **the user** can create the Metadata

So what does this mean?


IT MEANS THAT
PHYLOFORGE
NEEDS YOU
TO DO YOUR PART
Join the fight
and help make Science
a FAIRer place


Thank you
for your attention