



sharing and archiving research data

why, who, what, whom,
whose, when, where, how

Dr. Emily J. Kate



data steward



4.082



why?

journals/funders, reproducibility, science advancement
debt to future generations?

who?

you.

your supervisor will direct you, Emily can help

what?

“everything important”

raw/processed/meta data, results, code/protocols

whom?

colleagues, general public

whose?

ours, patients', cohort members, animal models...

when?

best: yesterday

second-best: now

pragmatic: when it's ready to publish

happens too often: never

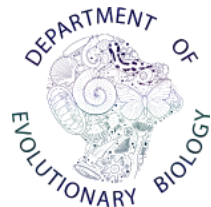
where?

“appropriate repository”

specific > general

institution > external

FAIRsharing, Re3data (or ask: supervisor, Emily, Niko?)



where? examples

- SRA (high-throughput sequencing)
 - ENA: European alternative
- GenBank (public DNA sequences)
- RSCB PDB (3D protein structures)
- EMDataResource (cryo-EM 3D prot. struct.)
- zenodo/PHAIDRA (literally everything)

→ overlaps, intersections!

how?

use **LICENSE**, persistent identifier, OSF

example

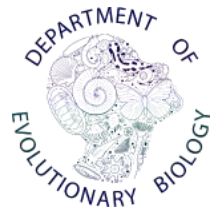
archiving and sharing
the *Pycnogonum litorale* genome*

* used for illustration purposes; please be a responsible adult when using this example



why?

(we can skip this)



who?*



Brenneis



Baranyi

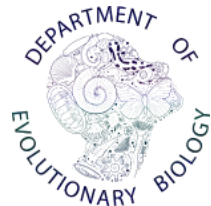


Papadopoulos



Wanninger

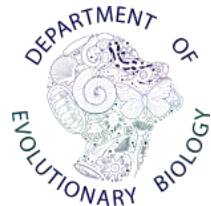
* possibly more, but you get the gist



what?*

- raw genomic reads (short, long)
- raw txomic reads (short, long)
- genome assembly
- genome annotation
- exp. protocols
- code

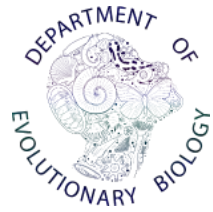
* again, an illustrative example, please don't take this literally



whom?*

- arthropod/chelicerate/pycnogonid researchers
- comp. geneticists
- evo-devo
- phylogeneticists
- rest of humanity

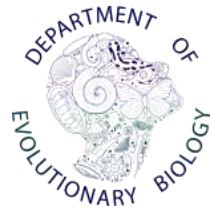
* not an exhaustive list; I'm trying to make a point here



whose?*

- Georg's
- the spiders'?

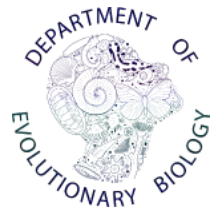
* ok I will stop the footnotes now



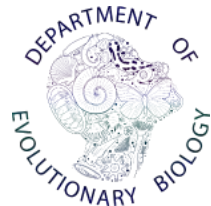
when?

publish draft & update? → can get scooped :(

↳ wait for paper/preprint

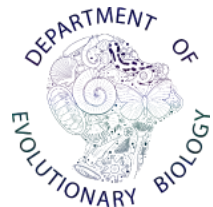


where?



where?

- raw gen. reads
(short, long)
 - raw txomic reads
(short, long)
 - genome assembly
 - interm. results
 - genome annotation
 - exp. protocols
 - code
-



where?

- raw gen. reads
(short, long)
- raw txomic reads
(short, long)
- genome assembly
- interm. results
- genome annotation
- exp. protocols
- code

during

 izoostorage

 /scratch/

 lab books

 GitHub



where?

- raw gen. reads (short, long)
- raw txomic reads (short, long)
- genome assembly
- interm. results
- genome annotation
- exp. protocols
- code

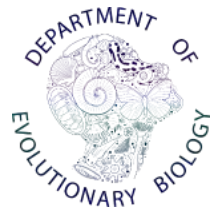
during

 izoostorage 

 /scratch/

 lab books

 GitHub 



where?

- raw gen. reads (short, long)
- raw txomic reads (short, long)
- genome assembly
- ~~interm. results~~
- genome annotation
- exp. protocols
- code

after

 izoostorage 

 SRA

 GenBank

 methods

 GitHub 

where?

- raw gen. reads (short, long)
- raw txomic reads (short, long)
- genome assembly
- ~~interm. results~~
- genome annotation
- exp. protocols
- code

after

 izoostorage 

 SRA 

 GenBank 

 methods 

 GitHub 

where?

- raw gen. reads (short, long)
- raw txomic reads (short, long)
- genome assembly
- ~~interm. results~~
- genome annotation
- exp. protocols
- code

after

 izoostorage 

 SRA   © TM

 GenBank   © TM

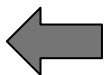
 methods   © TM

 GitHub   © TM

where?

- raw gen. reads (short, long)
- raw txomic reads (short, long)
- genome assembly
- ~~interm. results~~
- genome annotation
- exp. protocols
- code

if I really want to be
a model scientist:
zenodo/PHAIDRA



after

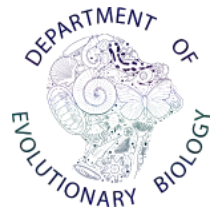
 izoostorage 

 SRA   © TM

 GenBank   © TM

 methods   © TM

 GitHub   © TM



where?

- raw gen. reads (short, long)
- raw txomic reads (short, long)
- genome assembly
- ~~interm. results~~
- genome annotation
- exp. protocols
- code

after

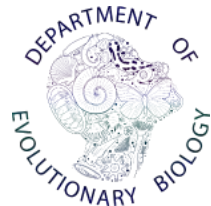
 izoostorage 

 SRA   © TM

 GenBank   © TM

 methods   © TM

 GitHub   © TM



extra tips from Emily (paraphrased)

- \$\$ for uploading can come from grants*
- please for the love of God use a license/DOI
- funders request RDM/sharing
- Open Science Framework → link outputs from same project

* if you think about it early enough

wh-questions

what
when
where
who
whom
which
whose
why
how

everything important
yesterday next best: now
somewhere easily reachable
you, really
all humanity?
(see “what”)
yours, test subjects’ / animals’
reproducibility, science advancement
DMP & follow-through!
