Joining data & communicating results with Quarto

CVEN 5837 - Summer 2023

Lars Schöbitz

https://cven5873-ss23.github.io/website/

Mortenson Center in Global Engineering
UNIVERSITY OF COLORADO BOULDER

https://cven5873-ss23.github.io/website/

Learning Objectives (for this week)

1. Learners can use Quarto and GitHub Pages to publish an HTML file.
2. Learners can add literature references to Quarto files using the navigation menu of RStudio visual editor.
3. Learners can cross-reference figures and tables within an Quarto file.
4. Learners can apply functions from the dplyr R Package to join multiple data sets.

Part 1: Joining data

We...

…have multiple data frames

…want to bring them together

```
professions <- read_csv(here::here("data/scientists/professions.csv"))
dates <- read_csv(here::here("data/scientists/dates.csv"))
works <- read_csv(here::here("scientists/works.csv"))
```

Data: Women in science

Information on 10 women in science who changed the world

name
Ada Lovelace
Marie Curie
Janaki Ammal
Chien-Shiung Wu
Katherine Johnson
Rosalind Franklin
Vera Rubin
Gladys West
Flossie Wong-Staal
Jennifer Doudna

Source: Discover Magazine

Inputs

professions | dates | works

name	profession
Ada Lovelace	Mathematician
Marie Curie	Physicist and Chemist
Janaki Ammal	Botanist
Chien-Shiung Wu	Physicist
Katherine Johnson	Mathematician
Rosalind Franklin	Chemist
Vera Rubin	Astronomer
Gladys West	Mathematician
Flossie Wong-Staal	Virologist and Molecular Biologist
Jennifer Doudna	Biochemist

Desired output

name	profession	birth_year	death_year	known_for
Ada Lovelace	Mathematician	NA	NA	first computer algorithm
Marie Curie	Physicist and Chemist	NA	NA	theory of radioactivity, discovery of elements polonium and radium, first woman to win a Nobel Prize
Janaki Ammal	Botanist	1897	1984	hybrid species, biodiversity protection
Chien-Shiung Wu	Physicist	1912	1997	confim and refine theory of radioactive beta decy, Wu experiment overturning theory of parity
Katherine Johnson	Mathematician	1918	2020	calculations of orbital mechanics critical to sending the first Americans into space
Rosalind Franklin	Chemist	1920	1958	NA

name	profession	birth_year	death_year	known_for
Vera Rubin	Astronomer	1928	2016	existence of dark matter
Gladys West	Mathematician	1930	NA	mathematical modeling of the shape of the Earth which served as the foundation of GPS technology
Flossie Wong-Staal	Virologist and Molecular Biologist	1947	NA	first scientist to clone HIV and create a map of its genes which led to a test for the virus
Jennifer Doudna	Biochemist	1964	NA	one of the primary developers of CRISPR, a ground-breaking technology for editing genomes

Inputs, reminder

```
1 names(professions)
```

```
[1] "name"       "profession"
```

```
1 names(dates)
```

```
[1] "name"       "birth_year"
"death_year"
```

```
1 names(works)
```

```
[1] "name"       "known_for"
```

```
1 nrow(professions)
```

```
[1] 10
```

```
1 nrow(dates)
```

```
[1] 8
```

```
1 nrow(works)
```

```
[1] 9
```

Joining data frames

Joining data frames

```
1 something_join(x, y)
```

- `left_join()`: all rows from x
- `right_join()`: all rows from y
- `full_join()`: all rows from both x and y
- ...

Setup

For the next few slides…

```r
x <- tibble(
  id = c(1, 2, 3),
  value_x = c("x1", "x2", "x3")
  )
```

```r
x
```

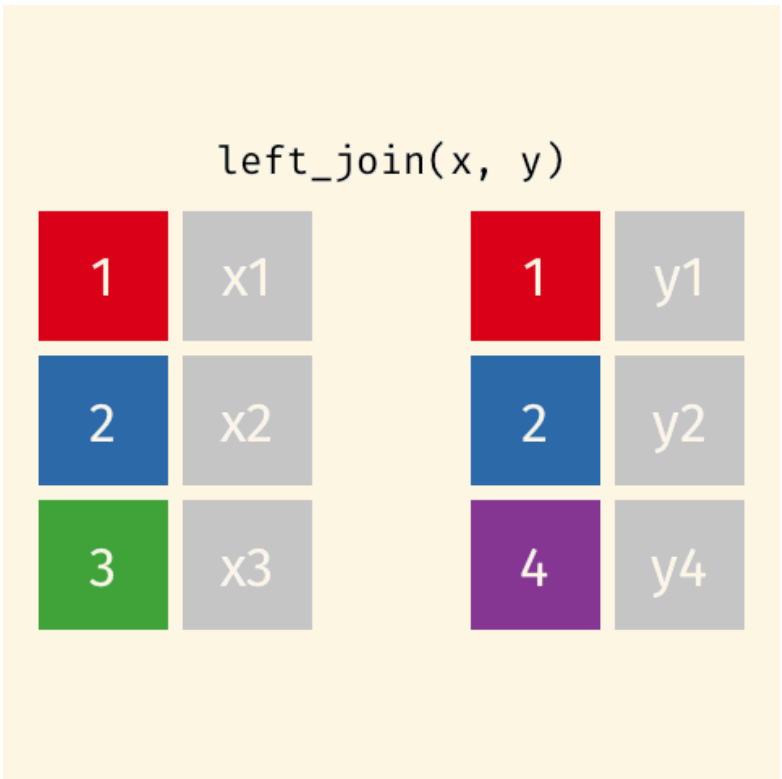
```
# A tibble: 3 × 2
     id value_x
  <dbl> <chr>
1     1 x1
2     2 x2
3     3 x3
```

```r
y <- tibble(
  id = c(1, 2, 4),
  value_y = c("y1", "y2", "y4")
  )
```

```r
y
```

```
# A tibble: 3 × 2
     id value_y
  <dbl> <chr>
1     1 y1
2     2 y2
3     4 y4
```

left_join()




```
1 left_join(x, y)
```

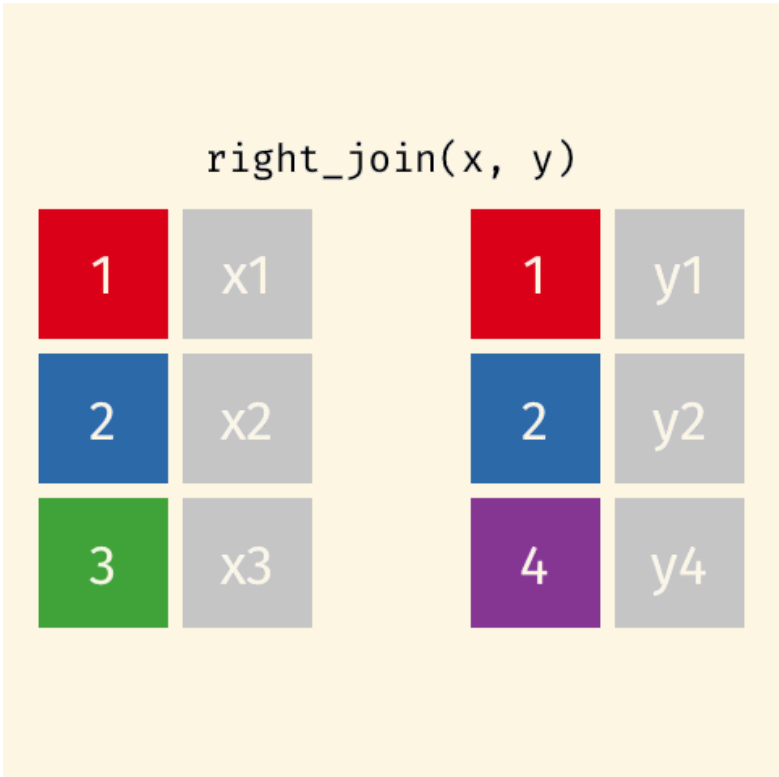
```
# A tibble: 3 × 3
     id value_x value_y
  <dbl> <chr>   <chr>
1     1 x1      y1
2     2 x2      y2
3     3 x3      <NA>
```

left_join()

```
professions %>%
  left_join(dates)
```

name	profession	birth_year	death_year
Ada Lovelace	Mathematician	NA	NA
Marie Curie	Physicist and Chemist	NA	NA
Janaki Ammal	Botanist	1897	1984
Chien-Shiung Wu	Physicist	1912	1997
Katherine Johnson	Mathematician	1918	2020
Rosalind Franklin	Chemist	1920	1958
Vera Rubin	Astronomer	1928	2016
Gladys West	Mathematician	1930	NA
Flossie Wong-Staal	Virologist and Molecular Biologist	1947	NA
Jennifer Doudna	Biochemist	1964	NA

right_join()




```
1 right_join(x, y)
```

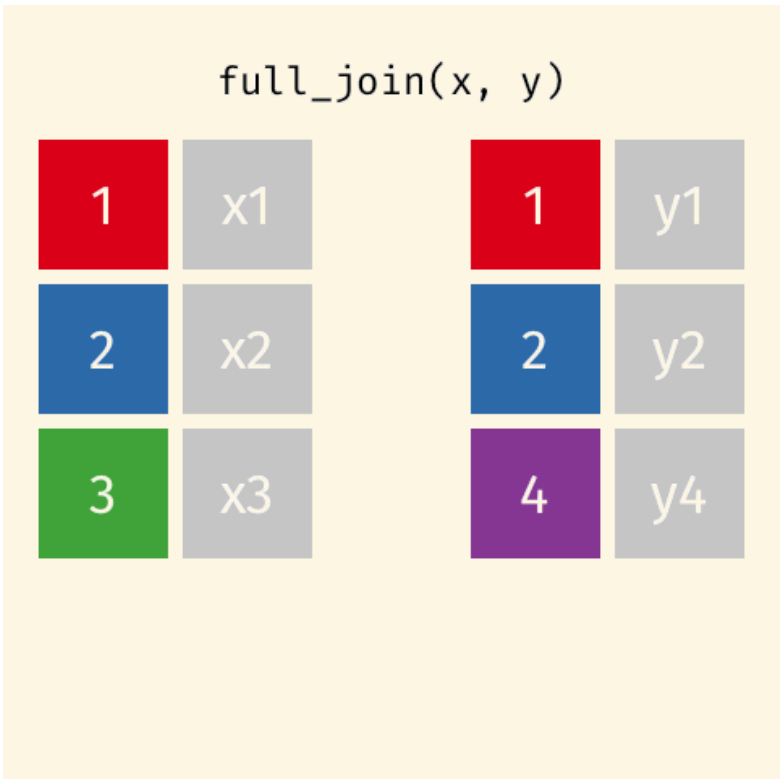
```
# A tibble: 3 × 3
     id value_x value_y
  <dbl> <chr>   <chr>
1     1 x1      y1
2     2 x2      y2
3     4 <NA>    y4
```

right_join()

```
professions %>%
  right_join(dates)
```

name	profession	birth_year	death_year
Janaki Ammal	Botanist	1897	1984
Chien-Shiung Wu	Physicist	1912	1997
Katherine Johnson	Mathematician	1918	2020
Rosalind Franklin	Chemist	1920	1958
Vera Rubin	Astronomer	1928	2016
Gladys West	Mathematician	1930	NA
Flossie Wong-Staal	Virologist and Molecular Biologist	1947	NA
Jennifer Doudna	Biochemist	1964	NA

full_join()




```
1 full_join(x, y)
```

```
# A tibble: 4 × 3
     id value_x value_y
  <dbl> <chr>   <chr>
1     1 x1      y1
2     2 x2      y2
3     3 x3      <NA>
4     4 <NA>    y4
```

full_join()

```
dates %>%
  full_join(works)
```

name	birth_year	death_year	known_for
Janaki Ammal	1897	1984	hybrid species, biodiversity protection
Chien-Shiung Wu	1912	1997	confim and refine theory of radioactive beta decy, Wu experiment overturning theory of parity
Katherine Johnson	1918	2020	calculations of orbital mechanics critical to sending the first Americans into space
Rosalind Franklin	1920	1958	NA
Vera Rubin	1928	2016	existence of dark matter
Gladys West	1930	NA	mathematical modeling of the shape of the Earth which served as the foundation of GPS technology

name	birth_year	death_year	known_for
Flossie Wong-Staal	1947	NA	first scientist to clone HIV and create a map of its genes which led to a test for the virus
Jennifer Doudna	1964	NA	one of the primary developers of CRISPR, a ground-breaking technology for editing genomes
Ada Lovelace	NA	NA	first computer algorithm
Marie Curie	NA	NA	theory of radioactivity, discovery of elements polonium and radium, first woman to win a Nobel Prize

Putting it altogether

```
professions %>%
  left_join(dates) %>%
  left_join(works) |>
  kable()
```

name	profession	birth_year	death_year	known_for
Ada Lovelace	Mathematician	NA	NA	first computer algorithm
Marie Curie	Physicist and Chemist	NA	NA	theory of radioactivity, discovery of elements polonium and radium, first woman to win a Nobel Prize
Janaki Ammal	Botanist	1897	1984	hybrid species, biodiversity protection
Chien-Shiung Wu	Physicist	1912	1997	confim and refine theory of radioactive beta decy, Wu experiment overturning theory of parity
Katherine Johnson	Mathematician	1918	2020	calculations of orbital mechanics critical to sending the first Americans into space

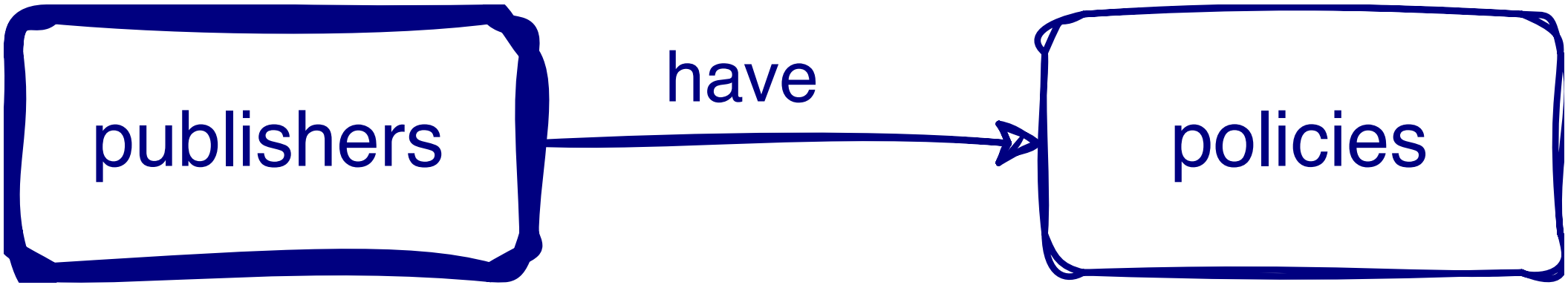
name	profession	birth_year	death_year	known_for
Rosalind Franklin	Chemist	1920	1958	NA
Vera Rubin	Astronomer	1928	2016	existence of dark matter
Gladys West	Mathematician	1930	NA	mathematical modeling of the shape of the Earth which served as the foundation of GPS technology
Flossie Wong-Staal	Virologist and Molecular Biologist	1947	NA	first scientist to clone HIV and create a map of its genes which led to a test for the virus
Jennifer Doudna	Biochemist	1964	NA	one of the primary developers of CRISPR, a ground-breaking technology for editing genomes

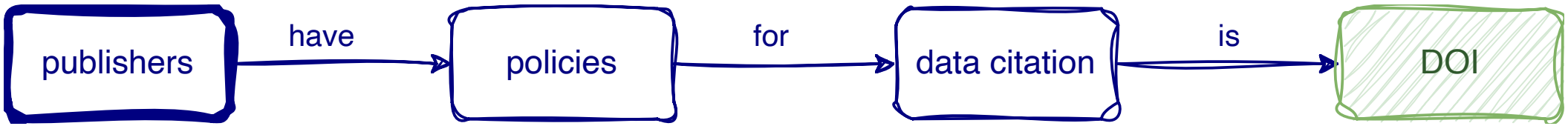
Part 2: Communicate results with Quarto

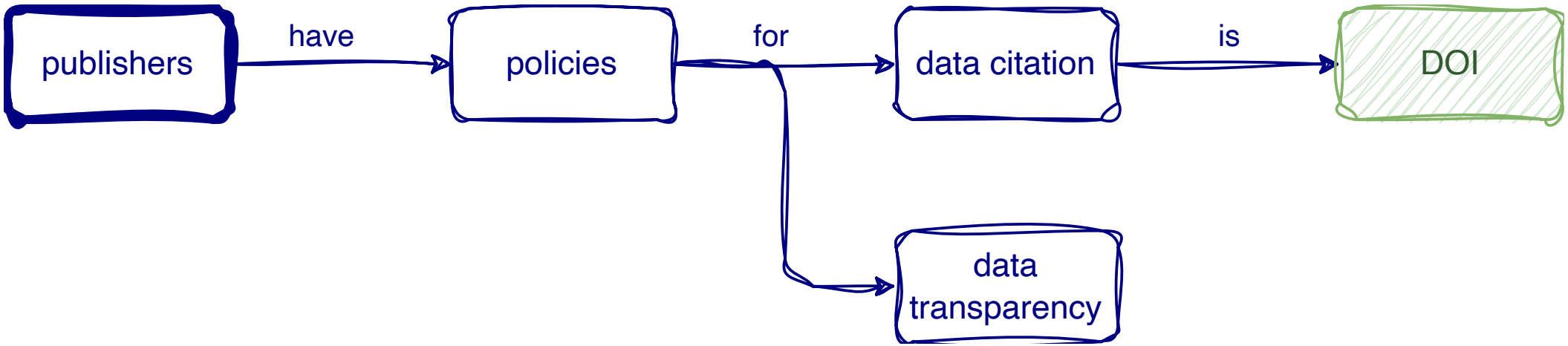
https://cven5873-ss23.github.io/website/

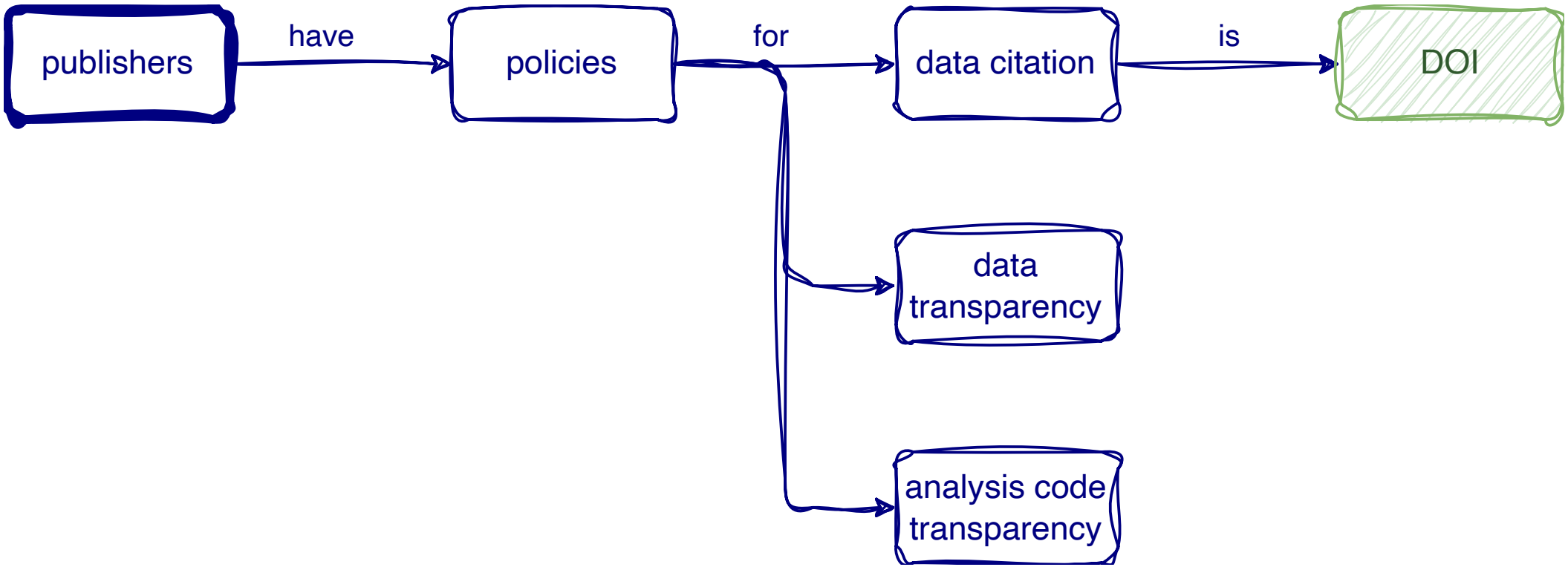
Why Quarto? -> Open Science

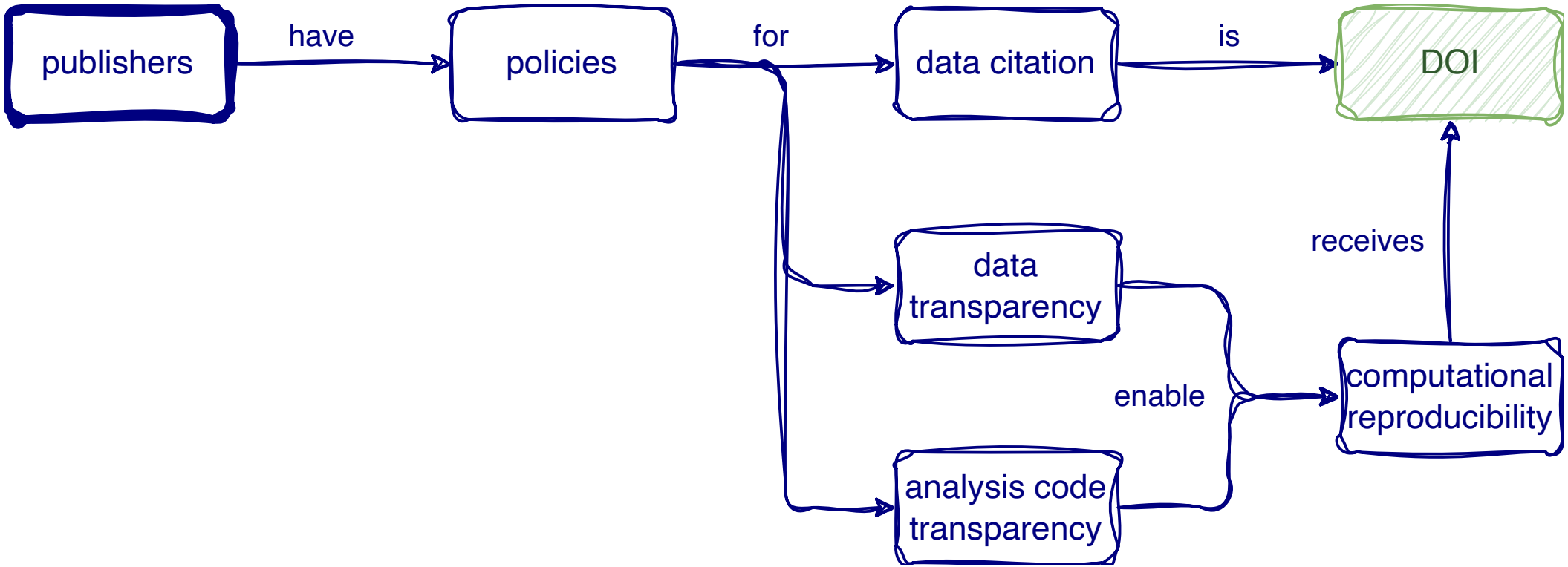
https://cven5873-ss23.github.io/website/

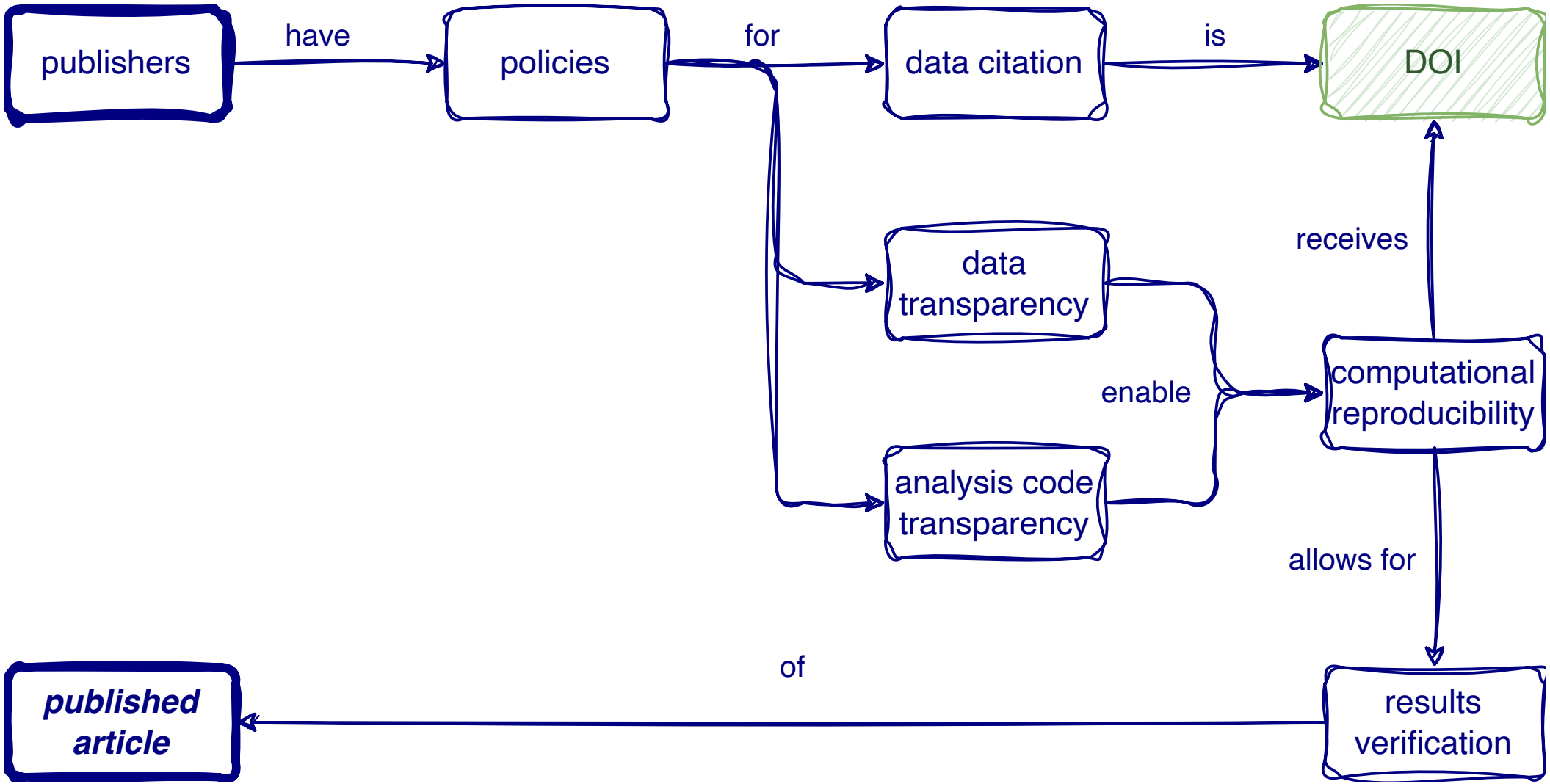

publishers
have
policies

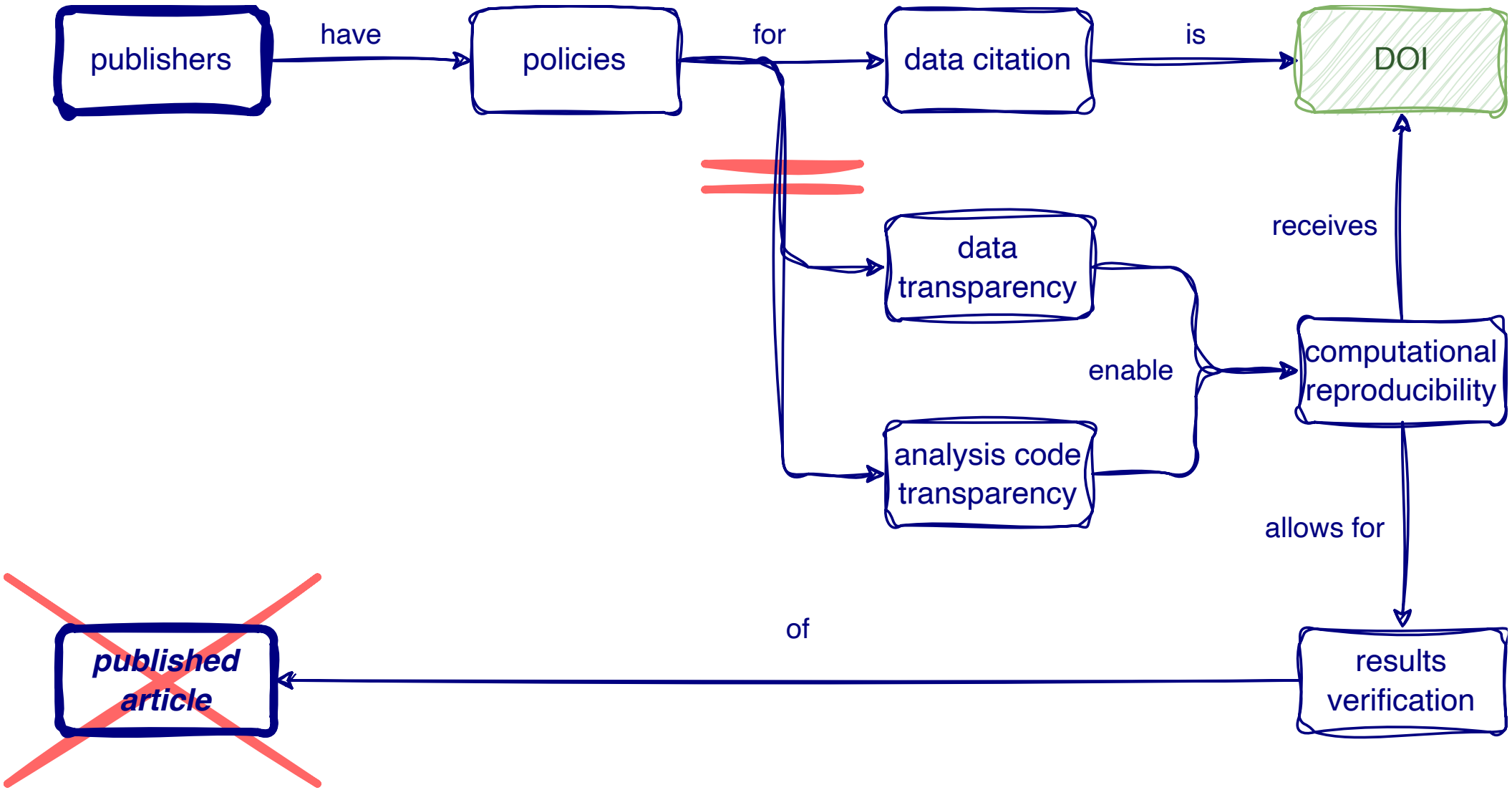

publishers
have
policies
for
data citation
is
DOI


publishers
have
policies
for
data citation
is
DOI
data transparency


publishers
have
policies
for
data citation
is
DOI
data transparency
analysis code transparency


publishers
have
policies
for
data citation
is
DOI
data transparency
receives
enable
computational reproducibility
analysis code transparency


publishers
have
policies
for
data citation
is
DOI
data transparency
receives
enable
computational reproducibility
analysis code transparency
allows for
of
published article
results verification

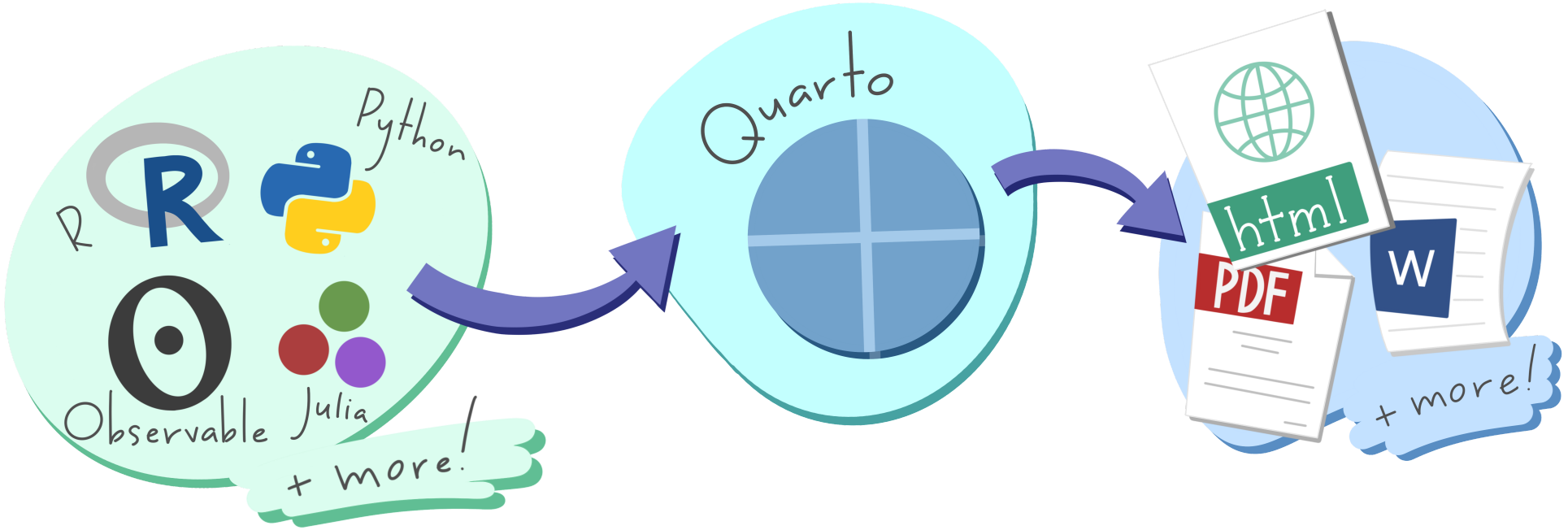

publishers
have
policies
for
data citation
is
DOI
data transparency
receives
enable
computational reproducibility
analysis code transparency
allows for
published article
of
results verification

What is Quarto?

https://cven5873-ss23.github.io/website/

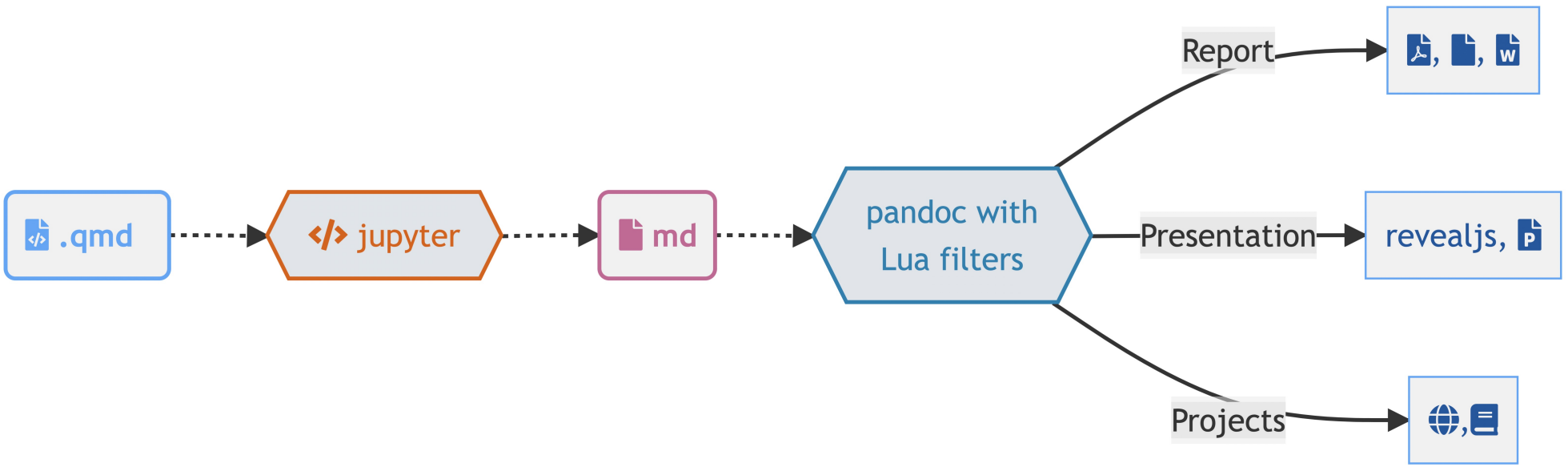
Quarto is a new, open-source, scientific and technical publishing system

the goal is to make the process of creating
and collaborating dramatically better



https://cven5873-ss23.github.io/website/

Quarto for literate programming



https://cven5873-ss23.github.io/website/

What is a `.qmd`?

> A Quarto document i.e. a `.qmd` is a plain text file

Metadata (YAML)

```yaml
format: html
engine: knitr
```

```yaml
format: html
engine: jupyter
```

Code

````
```{r}
library(dplyr)

mtcars |>
  dplyr::group_by(cyl) |>
  dplyr::summarize(mean = mean(mp
```
````

````
```{python}
from siuba import _, group_by, su
from siuba.data import mtcars
(mtcars
  >> group_by(_.cyl)
  >> summarize(avg_mpg = _.mpg.me
  )
```
````

Text

```
# Heading 1
This is a sentence with some **bold text**, some *italic text* and an
```

Quarto makes moving between formats straightforward

Document

lesson-1.qmd

```
1 title: "Lesson 1"
2 format: html
```

Presentation

lesson-1.qmd

```
1 title: "Lesson 1"
2 format: revealjs
```

Website

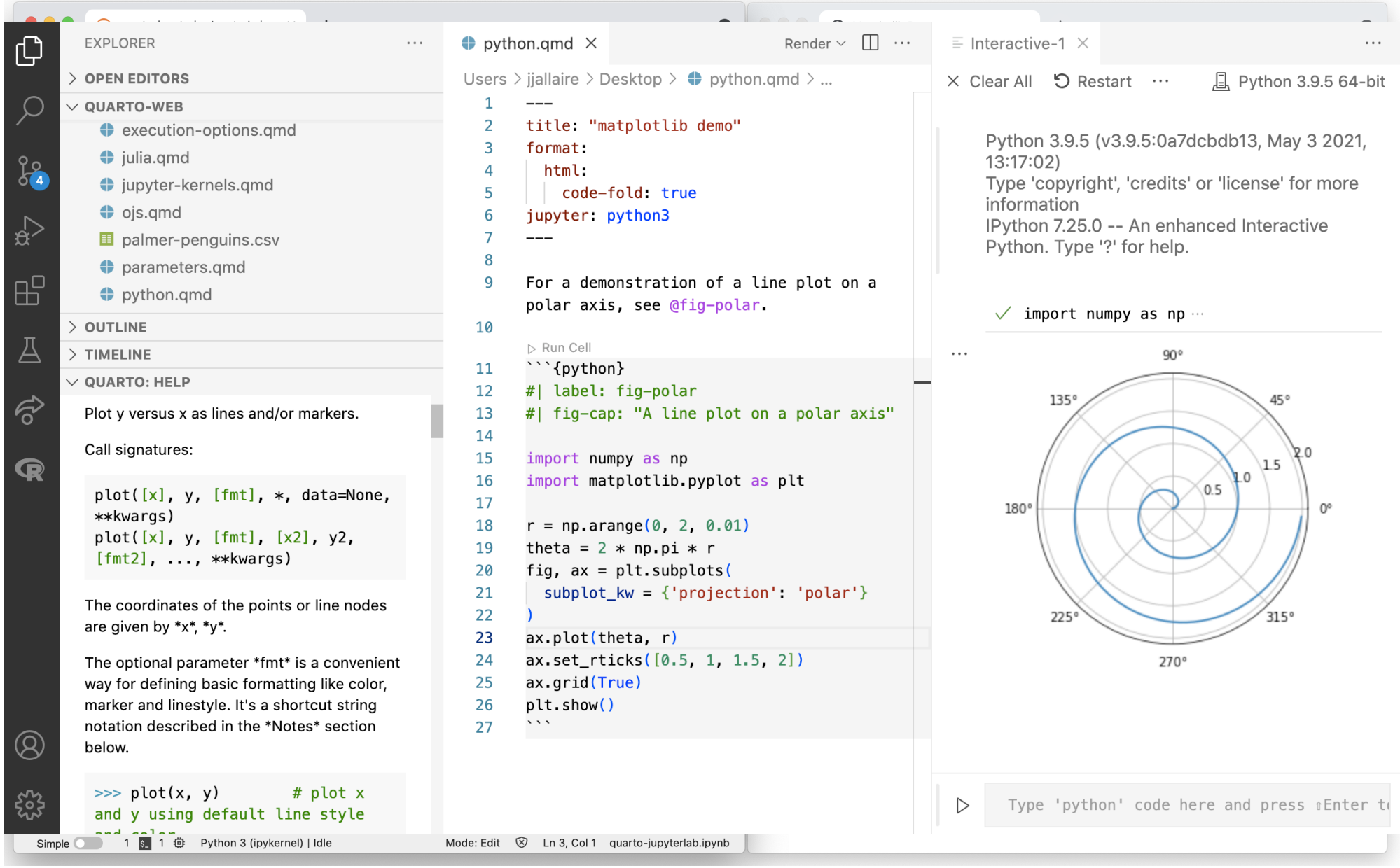
_quarto.yml

```
1 project:
2   type: website
3
4 website:
5   navbar:
6     left:
7       - lesson-1.qmd
```

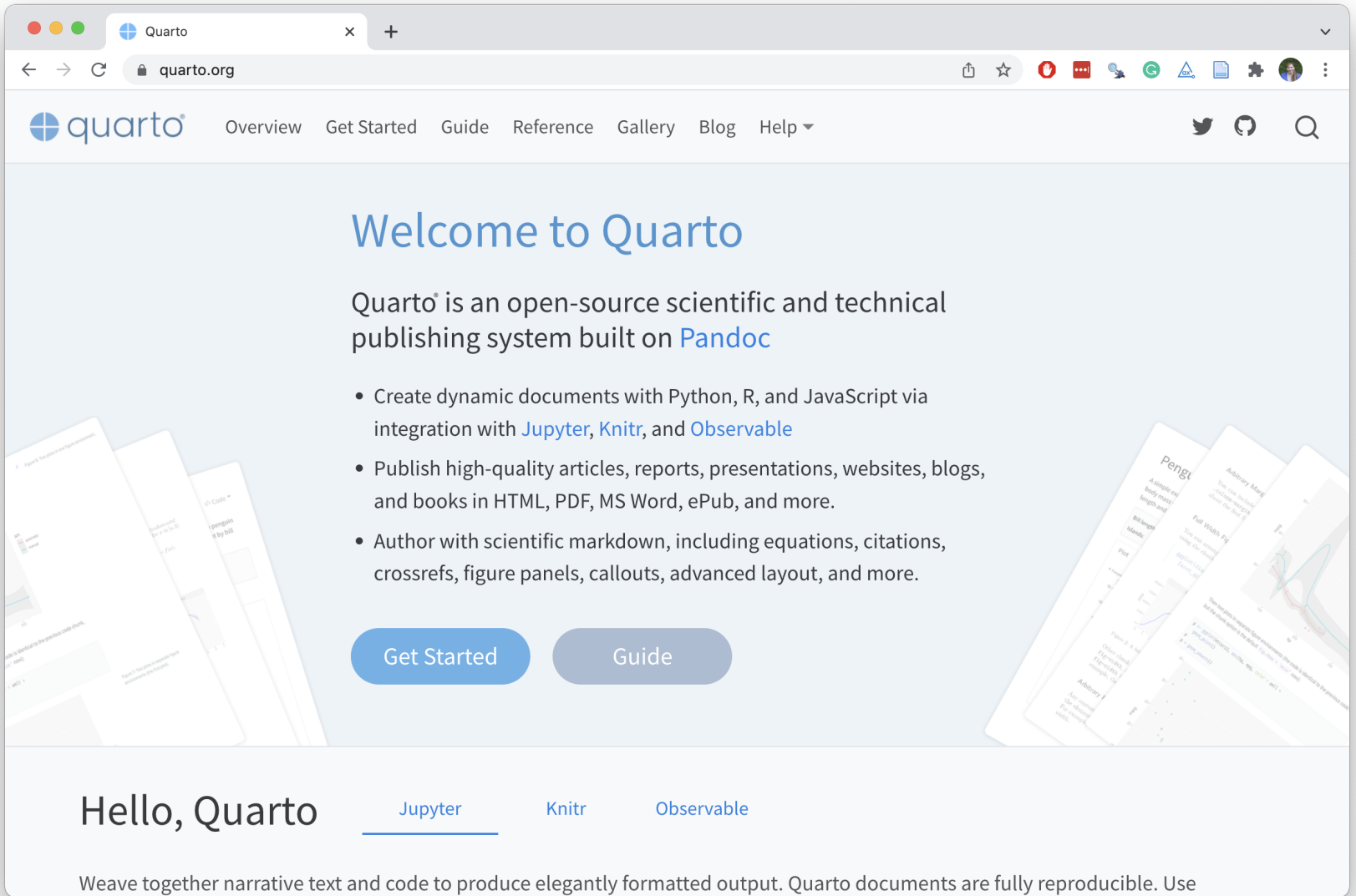
https://cven5873-ss23.github.io/website/

Comfort of your own workspace

https://cven5873-ss23.github.io/website/


EXPLORER
OPEN EDITORS
QUARTO-WEB
execution-options.qmd
julia.qmd
jupyter-kernels.qmd
ojs.qmd
palmer-penguins.csv
parameters.qmd
python.qmd
OUTLINE
TIMELINE
QUARTO: HELP
Plot y versus x as lines and/or markers.
Call signatures:
plot([x], y, [fmt], *, data=None, **kwargs)
plot([x], y, [fmt], [x2], y2, [fmt2], ..., **kwargs)
The coordinates of the points or line nodes are given by *x*, *y*.
The optional parameter *fmt* is a convenient way for defining basic formatting like color, marker and linestyle. It's a shortcut string notation described in the *Notes* section below.
>>> plot(x, y) # plot x and y using default line style
python.qmd
Render
Users > jjallaire > Desktop > python.qmd > ...
1 ---
2 title: "matplotlib demo"
3 format:
4 html:
5 code-fold: true
6 jupyter: python3
7 ---
8
9 For a demonstration of a line plot on a polar axis, see @fig-polar.
10
Run Cell
11 ```{python}
12 #| label: fig-polar
13 #| fig-cap: "A line plot on a polar axis"
14
15 import numpy as np
16 import matplotlib.pyplot as plt
17
18 r = np.arange(0, 2, 0.01)
19 theta = 2 * np.pi * r
20 fig, ax = plt.subplots(
21 subplot_kw = {'projection': 'polar'}
22)
23 ax.plot(theta, r)
24 ax.set_rticks([0.5, 1, 1.5, 2])
25 ax.grid(True)
26 plt.show()
27 ```
Interactive-1
Clear All
Restart
Python 3.9.5 64-bit
Python 3.9.5 (v3.9.5:0a7dcbdb13, May 3 2021, 13:17:02)
Type 'copyright', 'credits' or 'license' for more information
IPython 7.25.0 -- An enhanced Interactive Python. Type '?' for help.
import numpy as np
90°
135°
45°
180°
0°
225°
315°
270°
0.5
1.0
1.5
2.0
Type 'python' code here and press ⇧Enter t
Simple
1
1
Python 3 (ipykernel) | Idle
Mode: Edit
Ln 3, Col 1
quarto-jupyterlab.ipynb

Rich Documentation



https://cven5873-ss23.github.io/website/

Live Coding Exercise: Write a report

https://cven5873-ss23.github.io/website/

Clone GitHub repository from GitHub

1. Open the GitHub Organisation for the course: https://github.com/cven5873-ss23/
2. Locate the **wk-06** repository with your username `wk-06-GITHUB-USERNAME`
3. Follow along with me

Break

https://cven5873-ss23.github.io/website/

Cross-references

- no space between `{r}` and `#| tbl-cap: "A table"`
- spelling tbl not tab
- no spaces (use dashes in `label`)

Photo by Blake Wisz

See Table 1…

```
```{r}
#| tbl-cap: "A table"
#| label: tbl-simple-table

tibble(
  id = c(1, 2, 3),
  name = c("X", "Y", "Z")
) %>%
  knitr::kable()
```
```

Table 1: A table

id	name

id	name
1	X
2	Y
3	Z

Homework week 5

https://cven5873-ss23.github.io/website/

Homework due dates

- All material on course website
- Homework assignment & learning reflection due: **Friday, 14th July**

Capstone Project Report

https://cven5873-ss23.github.io/website/

Information

- All material shared on website by Monday, 17th July
- Due date for report: Friday, 28th July

https://cven5873-ss23.github.io/website/

Thanks! 

https://cven5873-ss23.github.io/website/

A large proportion of these slides are taken from Mine Çetinkaya Rundel's "Hello Quarto" presentation & Thomas Mock's "Quarto for the Curious" presentation

Slides created via revealjs and Quarto: https://quarto.org/docs/presentations/revealjs/ Access slides as PDF on GitHub