

WASH Canada - Module 5

Open Science Tools - authoring and publishing workflows for collaborative scientific writing

Lars Schöbitz

lschoebitz@ethz.ch

Global Health Engineering - ETH Zurich

May 25, 2023

larnsce.github.io/wash-canada/



Welcome! 🙌

Meet the lecturer

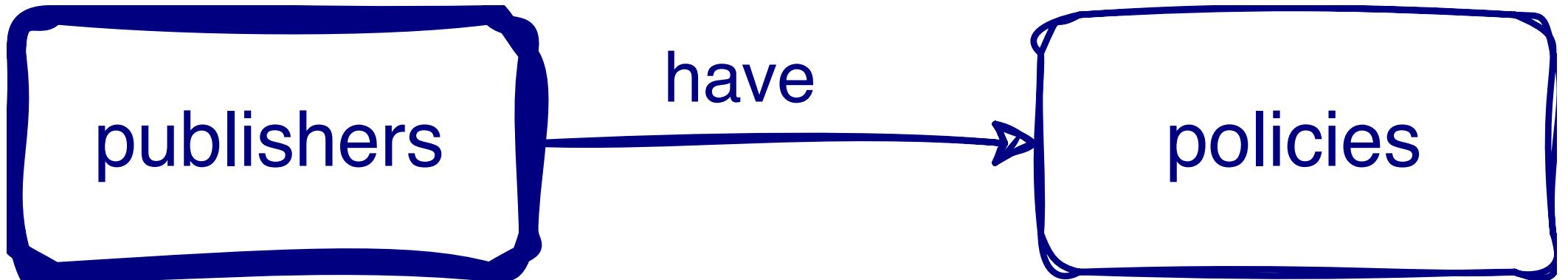
larnsce.github.io/wash-canada/

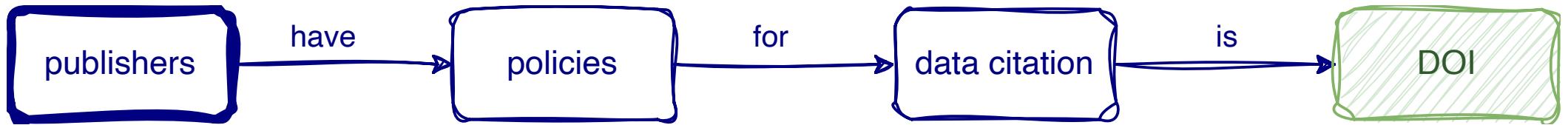
Lars Schöbitz (he/him)

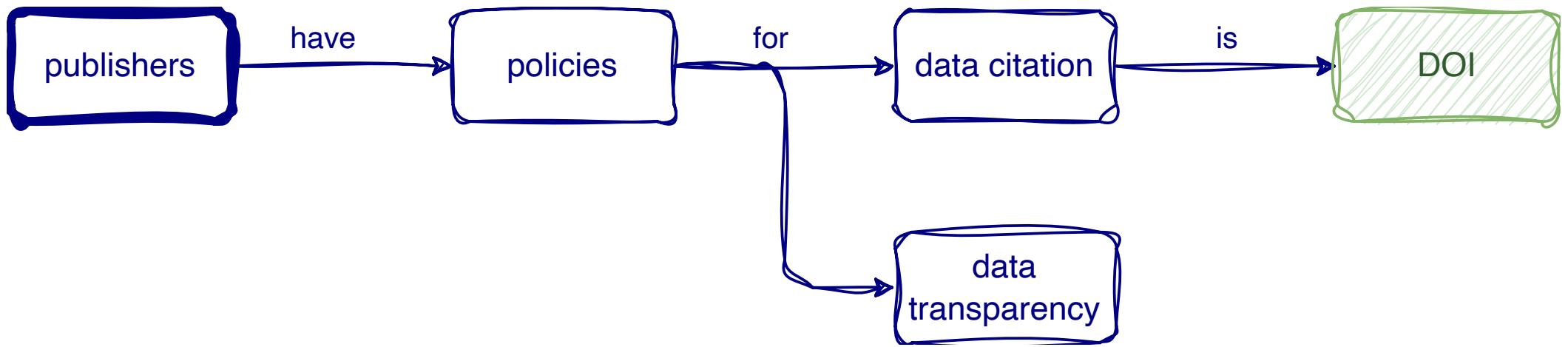


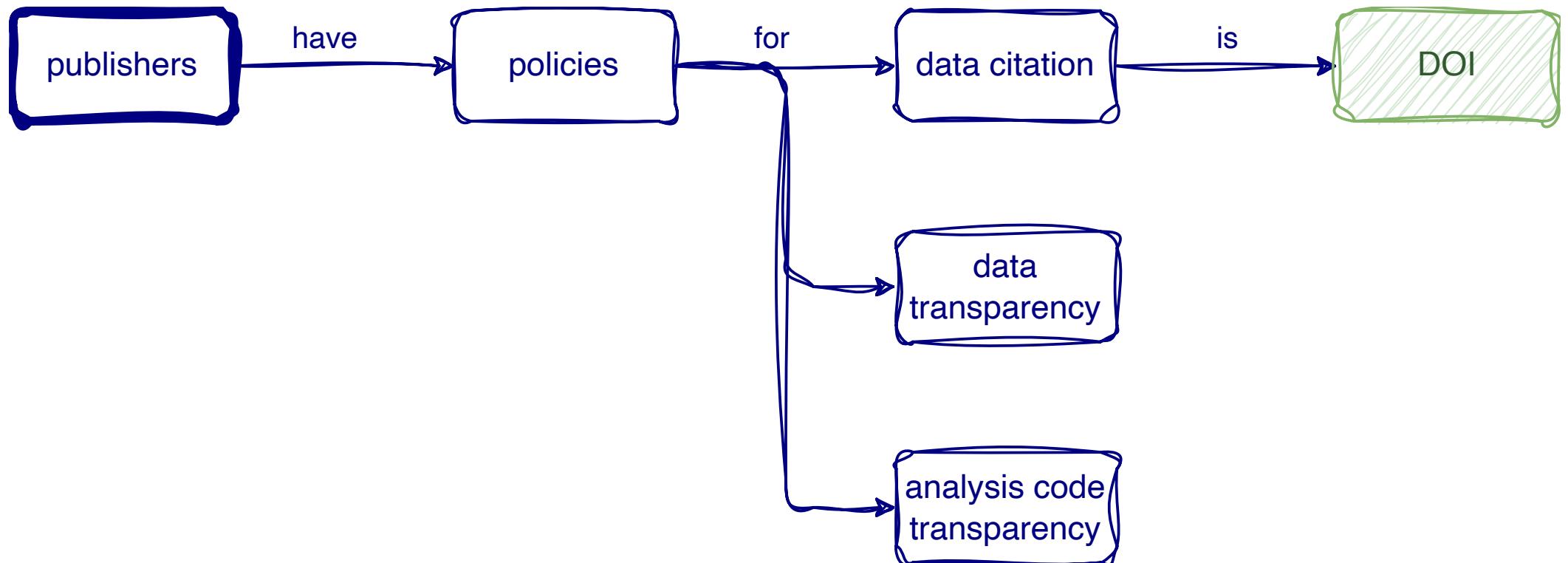
- Environmental Engineer
- Open Science Specialist at ETH Zurich
- Independent Instructor for Data Science with R

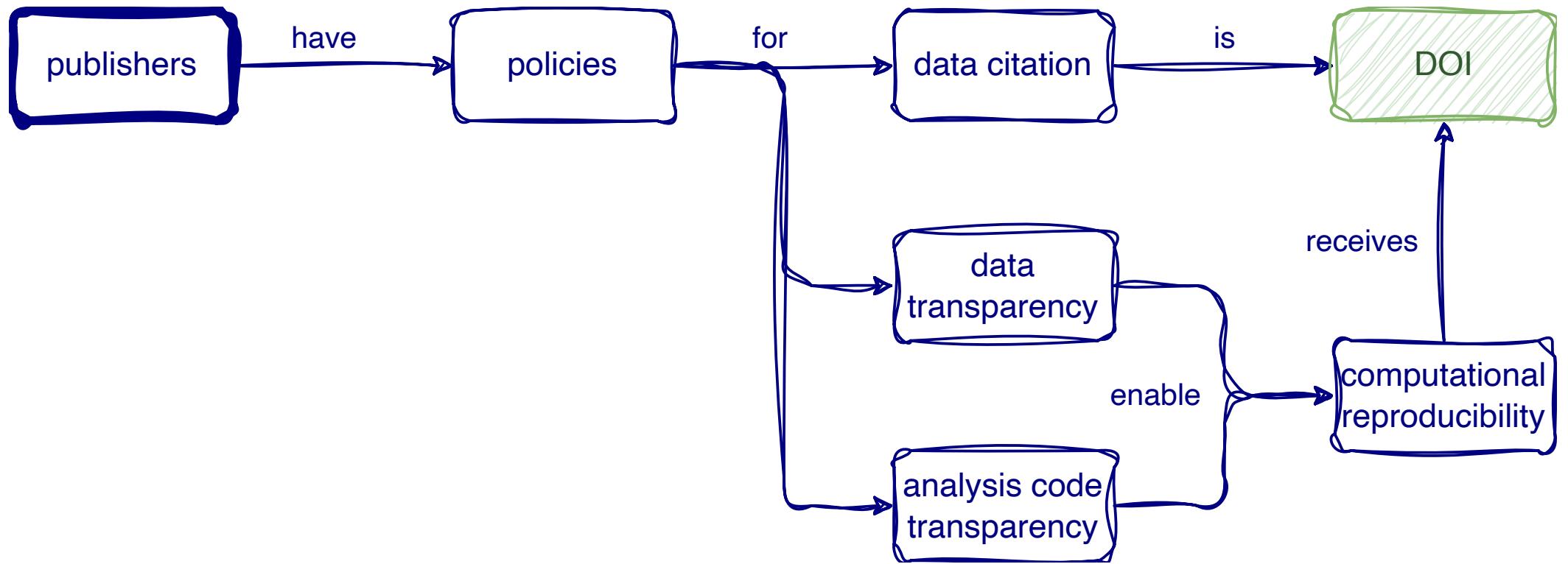
Why Open Science?

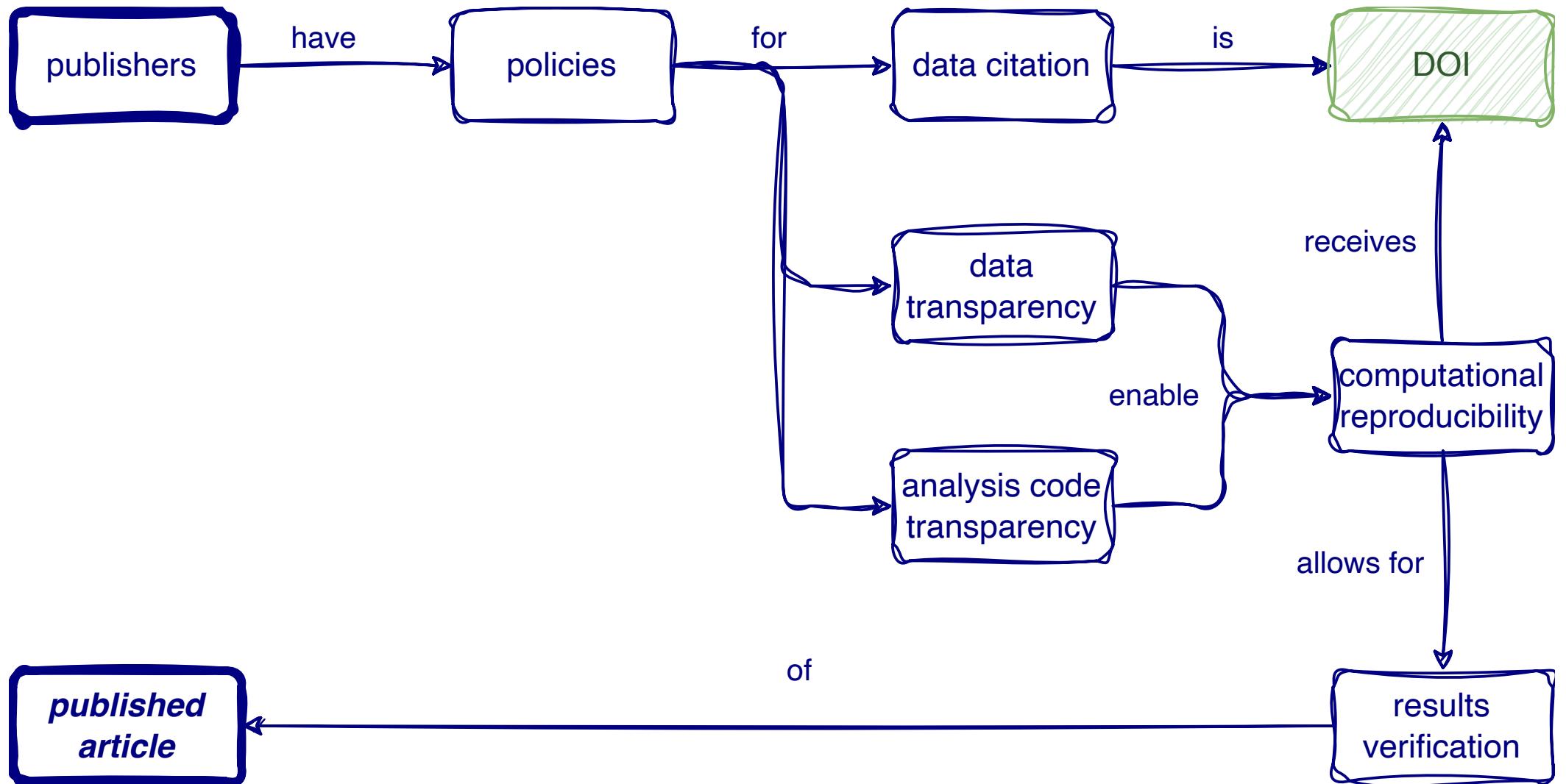


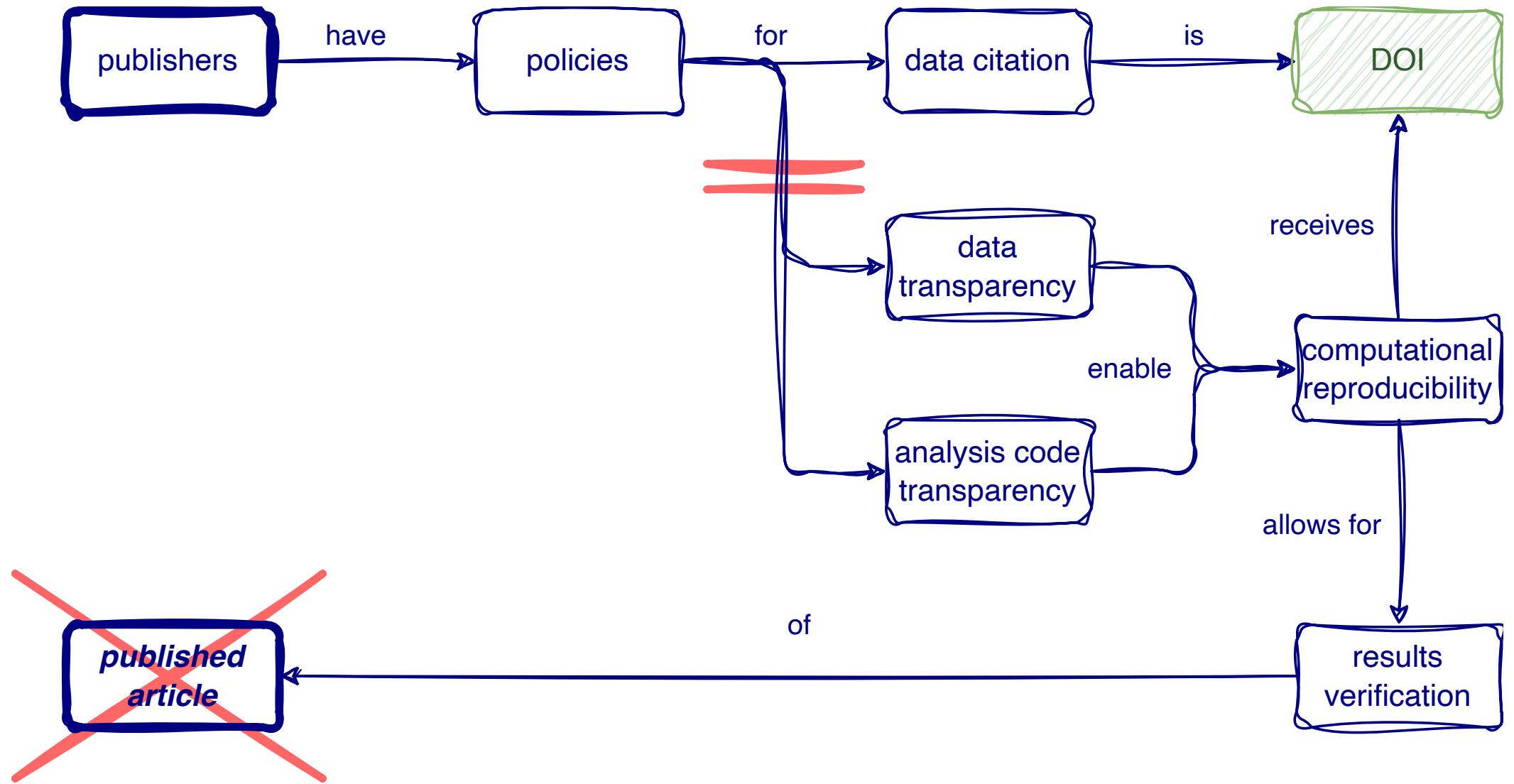








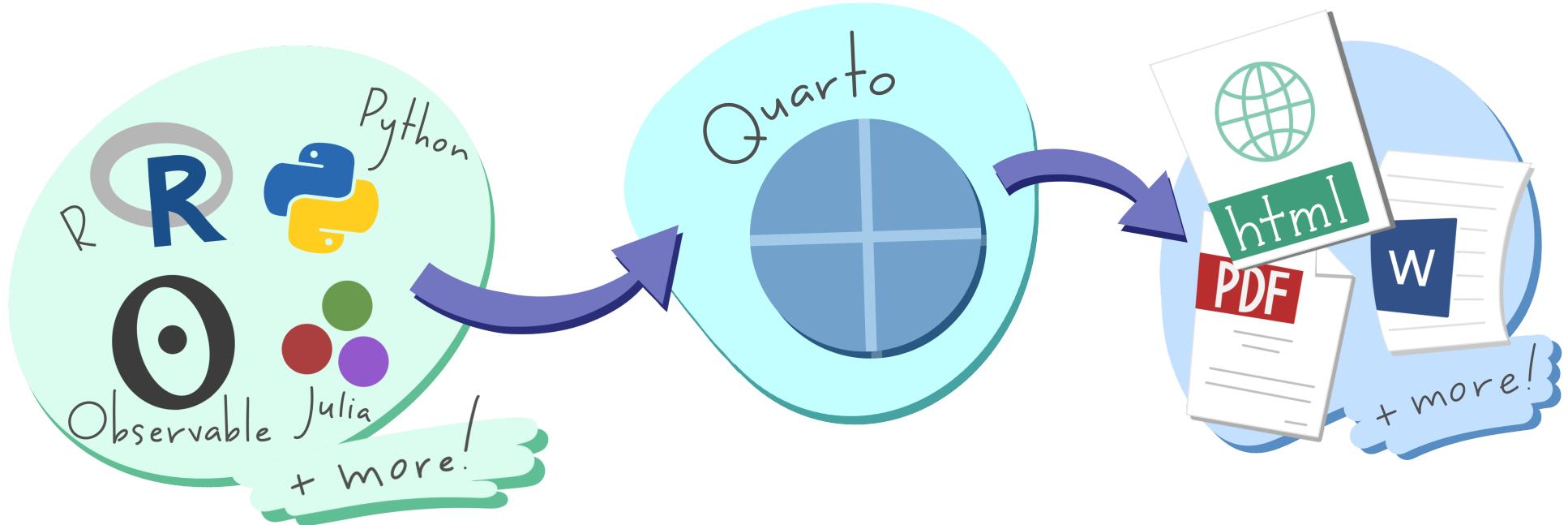




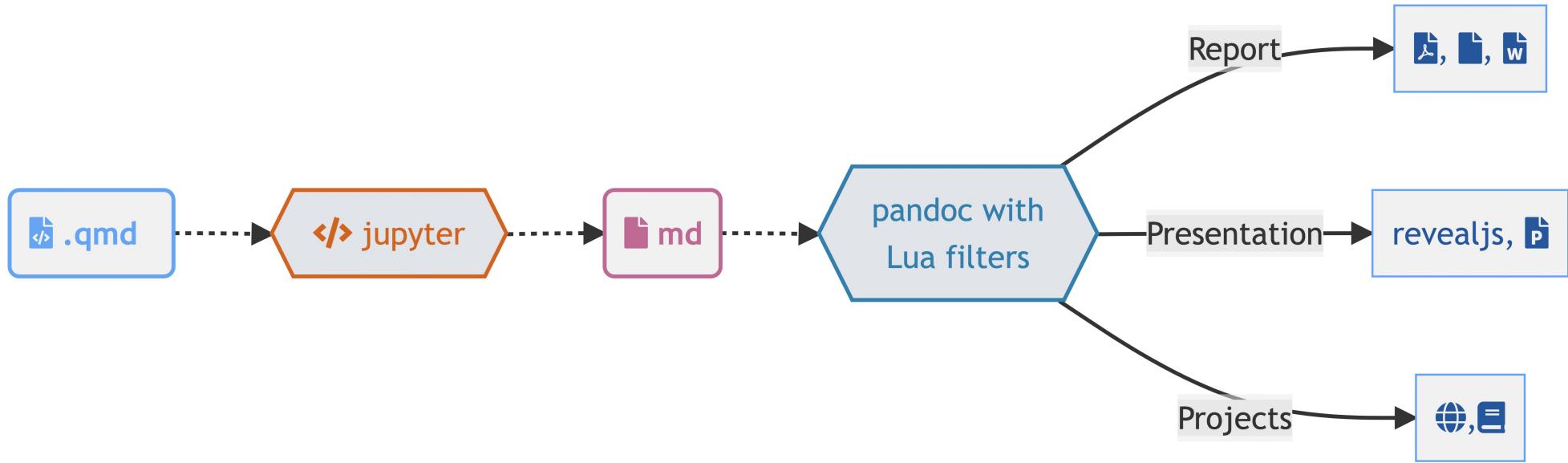
What is Quarto?

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

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



Quarto for literate programming



What is a `.qmd`?

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

Metadata (YAML)

```
1 format: html
2 engine: knitr
```

```
1 format: html
2 engine: jupyter
```

Code

```
1 ```{r}
2 library(dplyr)
3
4 mtcars |>
5   dplyr::group_by(cyl) |>
6   dplyr::summarize(mean = mean(mpg))
7 ````
```

```
1 ```{python}
2 from siuba import _, group_by, summarize
3 from siuba.data import mtcars
4 (mtcars
5   >> group_by(_.cyl)
6   >> summarize(avg_mpg = _.mpg.mean()))
7 )
8 ````
```

Text

```
1 # Heading 1
2 This is a sentence with some **bold text**, some *italic text* and an
3 ![image](image.png){fig-alt="Alt text for this image"}.
```

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
```

Comfort of your own workspace

The screenshot shows a Jupyter Notebook interface with the following components:

- Left Sidebar:** Includes icons for file operations, search, and help, along with sections for **OPEN EDITORS**, **QUARTO-WEB** (listing files like execution-options.qmd, julia.qmd, jupyter-kernels.qmd, ojs.qmd, palmer-penguins.csv, parameters.qmd, and python.qmd), **OUTLINE**, **TIMELINE**, and **QUARTO: HELP** (providing documentation for the plot function).
- Code Editor:** The active cell contains Python code for generating a polar plot. The code imports numpy and matplotlib.pyplot, defines a range of theta values, creates a polar subplot, and plots concentric circles.
- Terminal:** Shows a command-line interface with the Python 3 interpreter, displaying the generated plot.
- Plot Viewer:** A polar plot showing concentric circles centered at the origin. The plot has radial grid lines at 45° intervals and numerical labels (0.5, 1, 1.5, 2) along the outermost circle.
- Bottom Status Bar:** Displays "Simple" mode, line numbers (1, 2, 3), and the URL "larsce.github.io/wash-canada/".

larnsce.github.io/wash-canada/



Posit Cloud & RStudio IDE

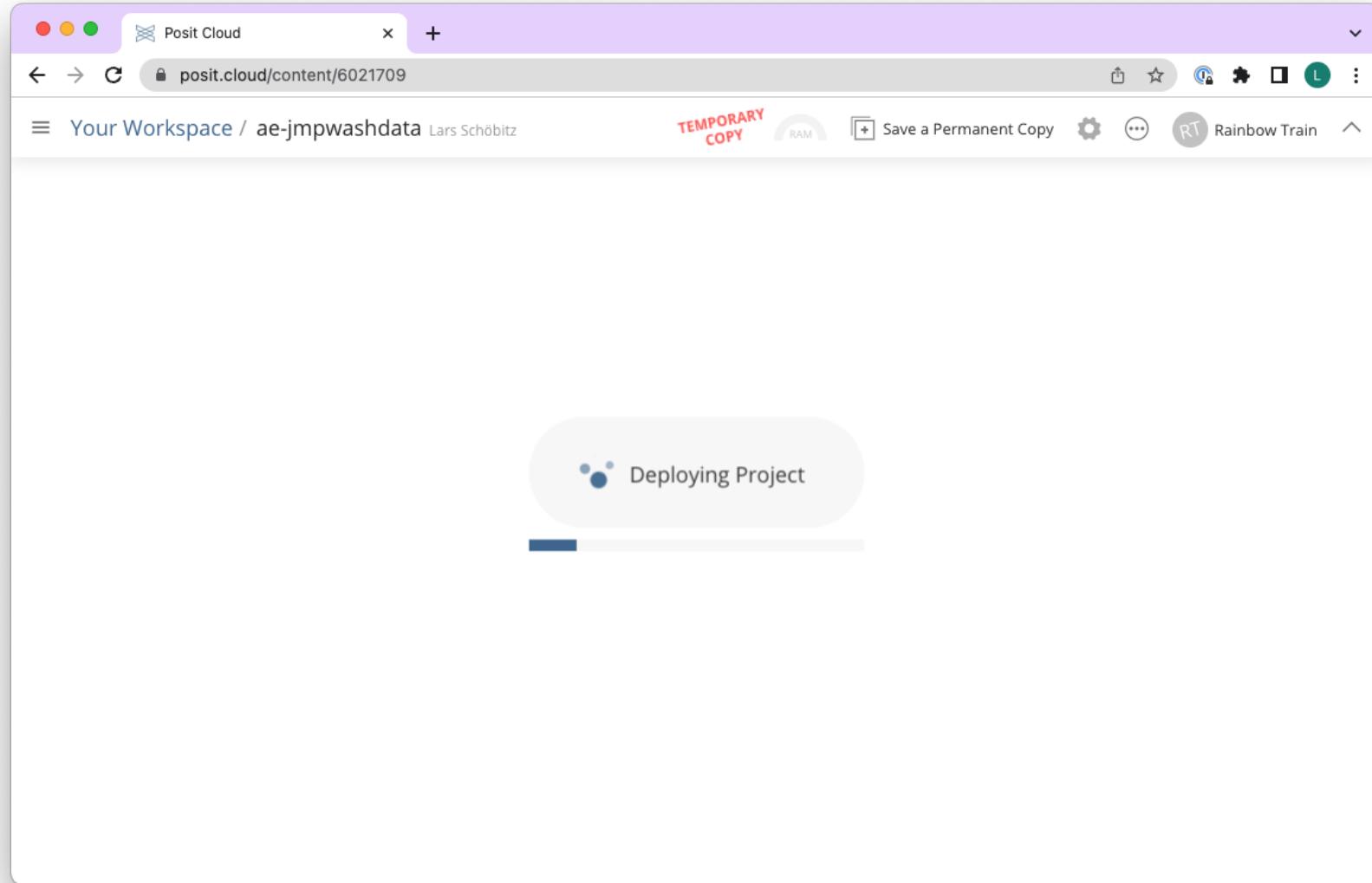
The screenshot shows the Posit Cloud web interface. At the top, there's a purple header bar with the Posit Cloud logo and a search bar containing the URL `posit.cloud/content/yourself?sort=name_asc`. Below the header, a navigation bar includes 'Your Workspace' (set to 'Rainbow Train'), 'Content' (which is selected and highlighted in blue), 'Usage', and 'About'. On the right, there's a user icon labeled 'Rainbow Train' with a dropdown arrow.

The main content area is titled 'Your Content (0)' and features a sidebar with three items: 'Your Content' (selected), 'Archive', and 'Trash'. To the right of the sidebar is a search and filter section with fields for 'TYPE', 'ACCESS', 'SORT', and a search icon. A message 'no content' is displayed below the search bar. In the bottom right corner of the content area, there's a 'New Project' button with a dropdown arrow.

At the very bottom of the page is a dark blue footer bar. It contains the 'posit' logo, social media icons for Facebook, Twitter, LinkedIn, and GitHub, and the text '© 2022 Posit Software, PBC'. There are also links for 'Terms' and 'Status'.

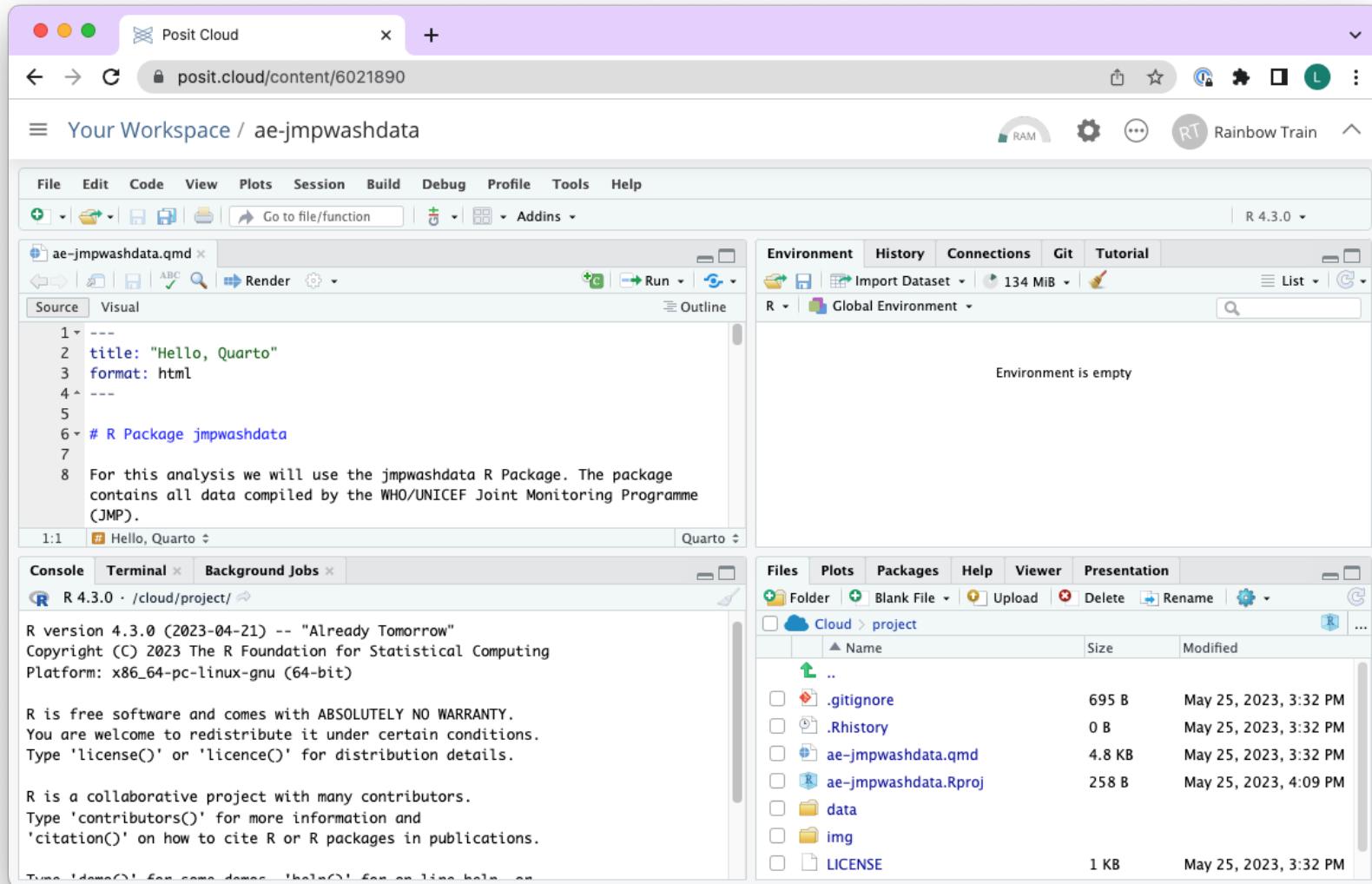
larnsce.github.io/wash-canada/

Posit Cloud & RStudio IDE



larnsce.github.io/wash-canada/

Posit Cloud & RStudio IDE



Let's say hello to Quarto

Open this link

<https://posit.cloud/content/6021709>

More resources for Quarto

- Video: [Reproducible authoring with Quarto](#)
- Video: [Openscapes: Hello Quarto with NASA Openscapes, RLadies Santa Barbara, and JJ Allaire](#)
- Quarto documentation: quarto.org
- Quarto gallery: quarto.org/docs/gallery
- Quarto tip a day: rstd.io/quartotip

openwashdata

larnsce.github.io/wash-canada/



openwashdata

- Receive **credit** for work that is not a scientific paper
- Give **recognition** to those that support your work
- Tell **stories** with data that haven't yet been told
- Meet **people** that care about data and code being open and reusable

The Opportunity

larnsce.github.io/wash-canada/



Journal Articles

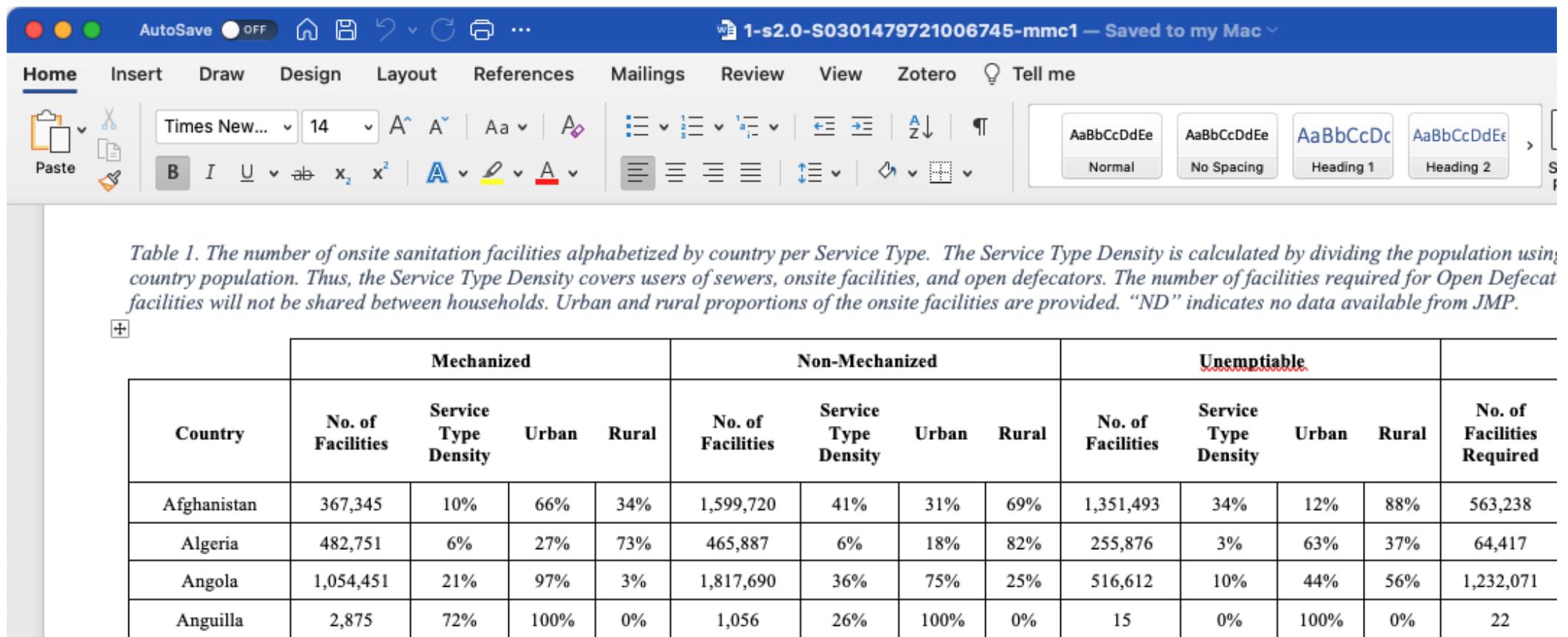
Appendix A. Supplementary data

The following is the supplementary data to this article:

 [Download : Download Word document \(152KB\)](#)

Multimedia component 1.

Journal Articles



The screenshot shows a Microsoft Word document window titled "1-s2.0-S0301479721006745-mmc1 — Saved to my Mac". The ribbon menu is visible at the top, showing Home, Insert, Draw, Design, Layout, References, Mailings, Review, View, Zotero, Tell me, and AutoSave (OFF). The Home tab is selected. Below the ribbon is the Word ribbon toolbar with various icons for paste, font, and paragraph styles.

Table 1. The number of onsite sanitation facilities alphabetized by country per Service Type. The Service Type Density is calculated by dividing the population using country population. Thus, the Service Type Density covers users of sewers, onsite facilities, and open defecators. The number of facilities required for Open Defecation facilities will not be shared between households. Urban and rural proportions of the onsite facilities are provided. "ND" indicates no data available from JMP.

Country	Mechanized				Non-Mechanized				Unemptiable				No. of Facilities Required
	No. of Facilities	Service Type Density	Urban	Rural	No. of Facilities	Service Type Density	Urban	Rural	No. of Facilities	Service Type Density	Urban	Rural	
Afghanistan	367,345	10%	66%	34%	1,599,720	41%	31%	69%	1,351,493	34%	12%	88%	563,238
Algeria	482,751	6%	27%	73%	465,887	6%	18%	82%	255,876	3%	63%	37%	64,417
Angola	1,054,451	21%	97%	3%	1,817,690	36%	75%	25%	516,612	10%	44%	56%	1,232,071
Anguilla	2,875	72%	100%	0%	1,056	26%	100%	0%	15	0%	100%	0%	22

The Journey

larnsce.github.io/wash-canada/

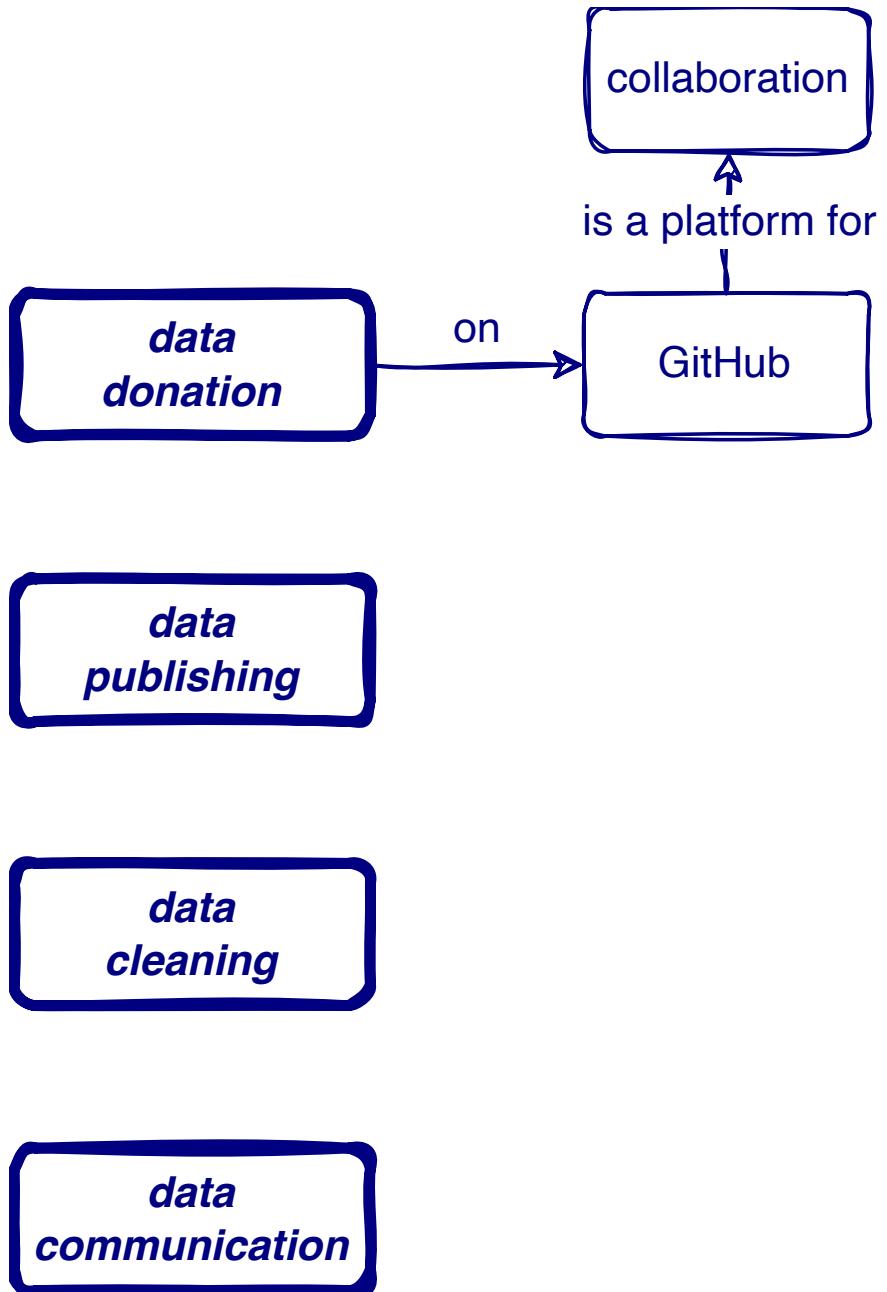


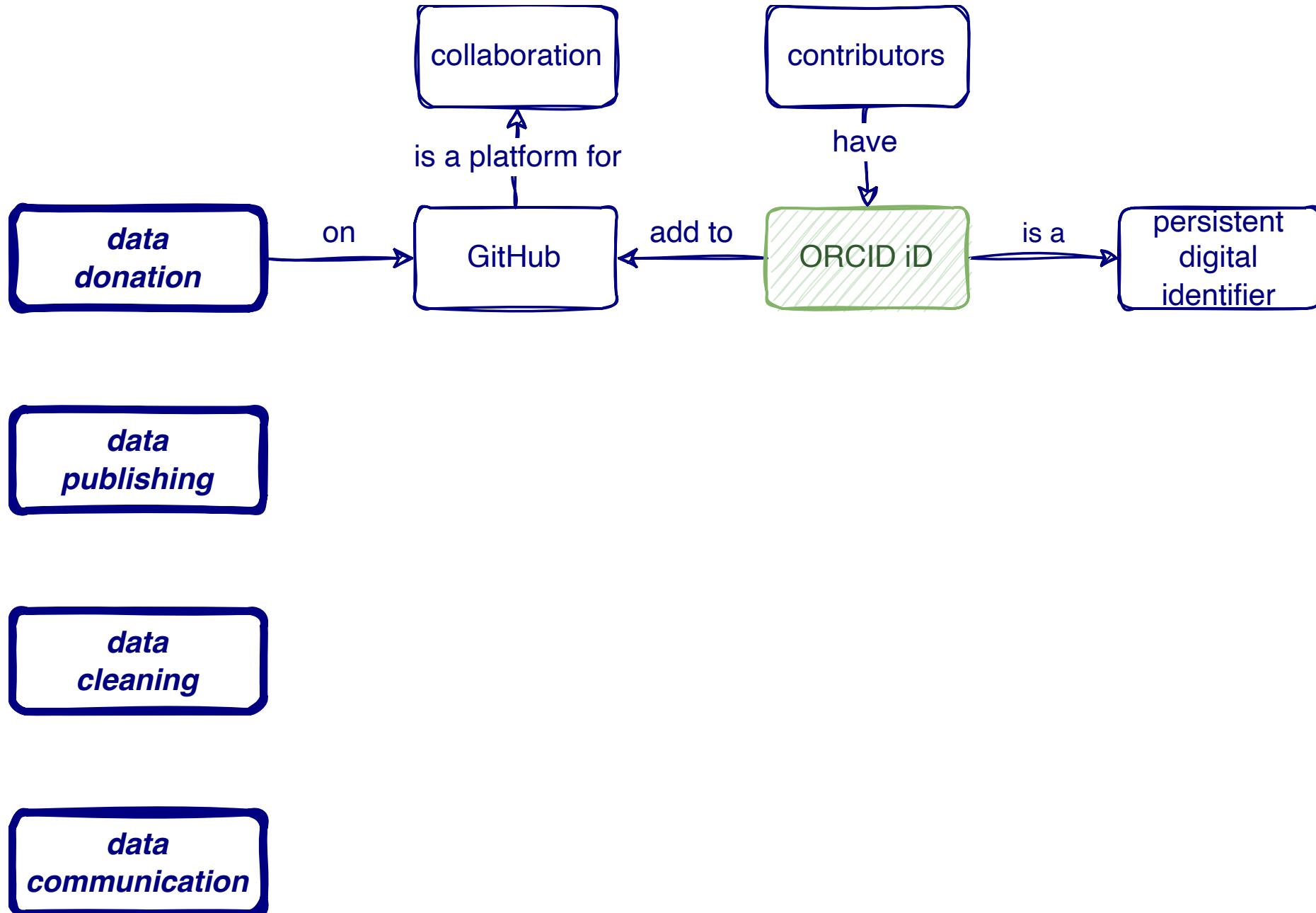
*data
donation*

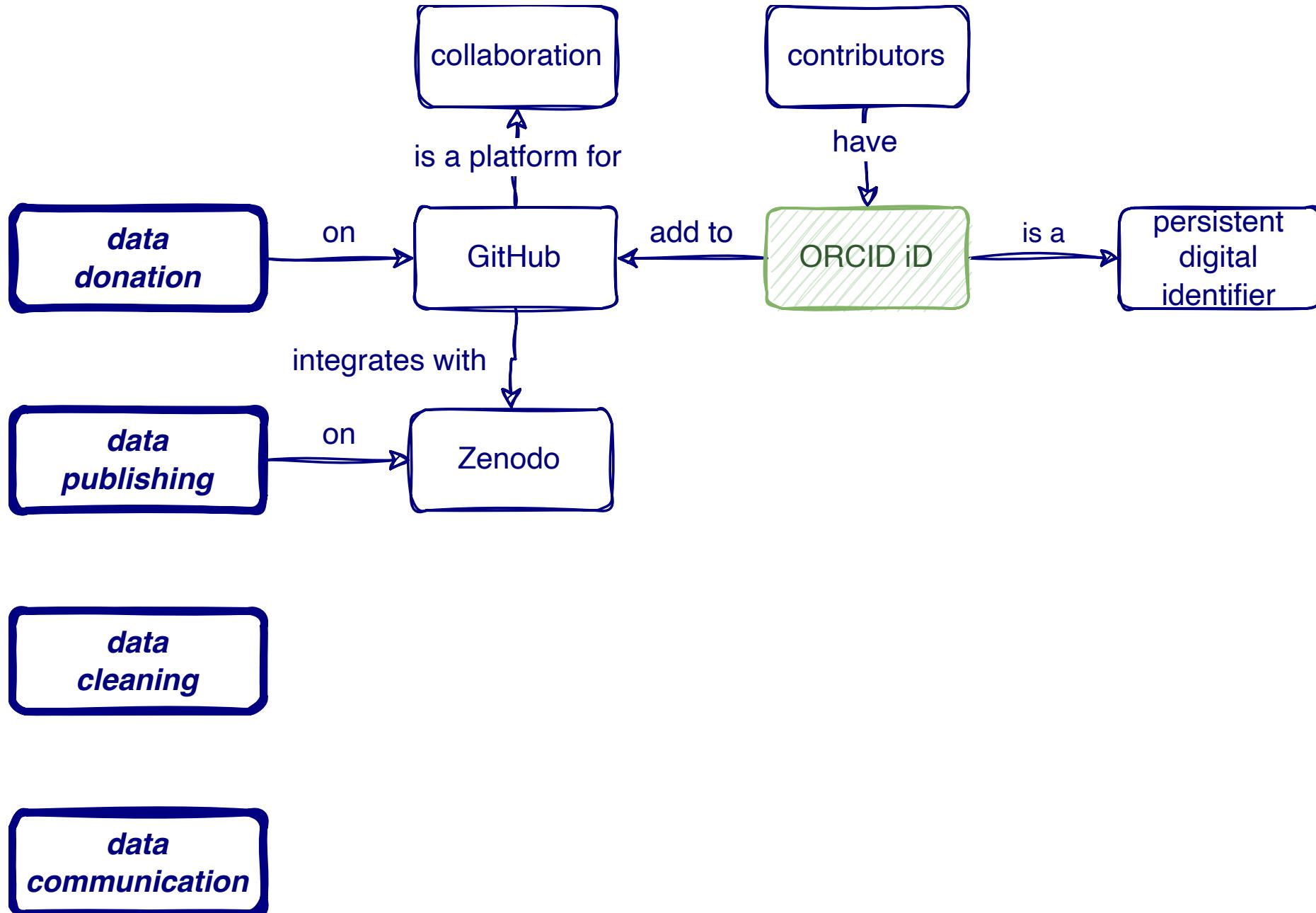
*data
publishing*

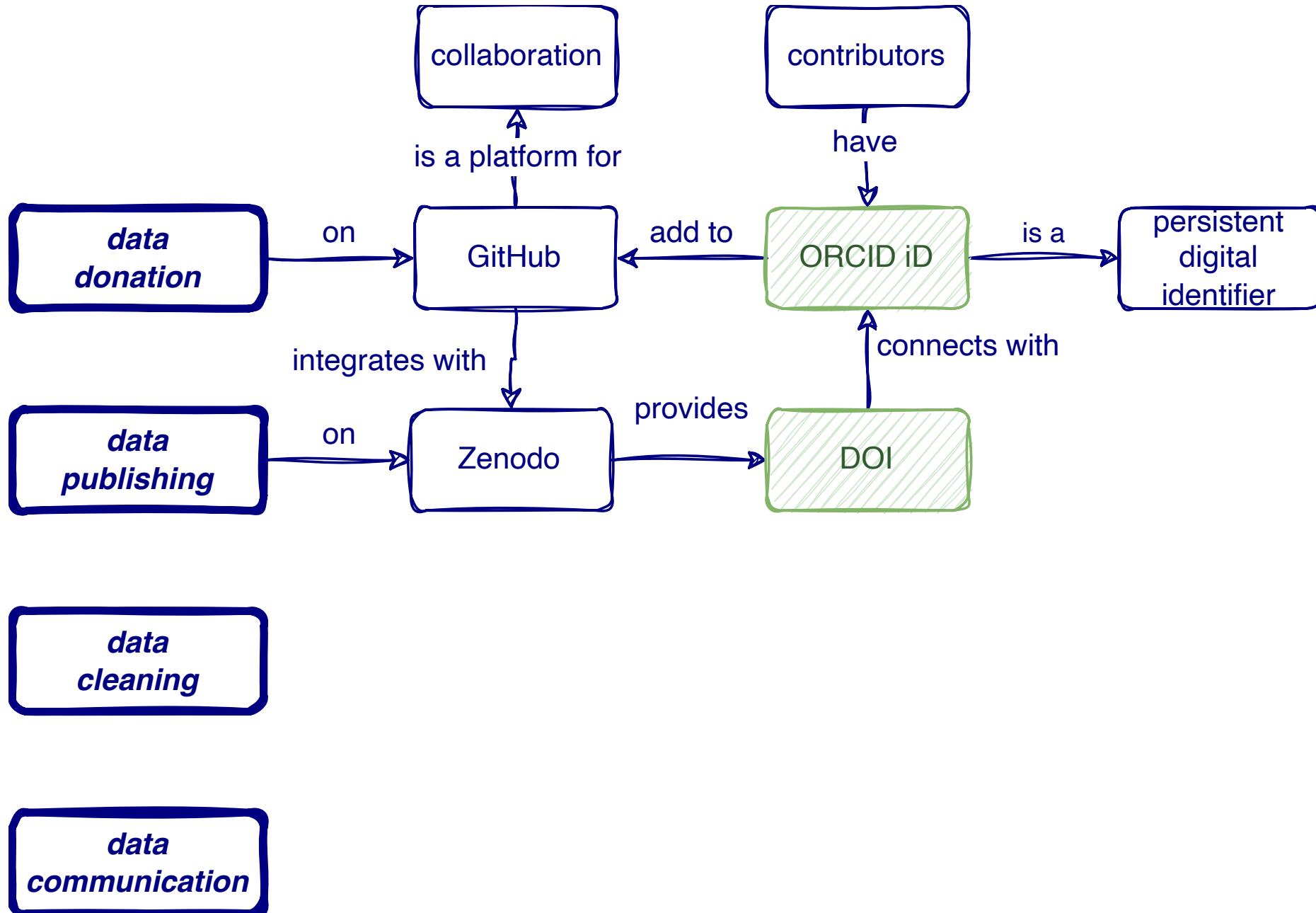
*data
cleaning*

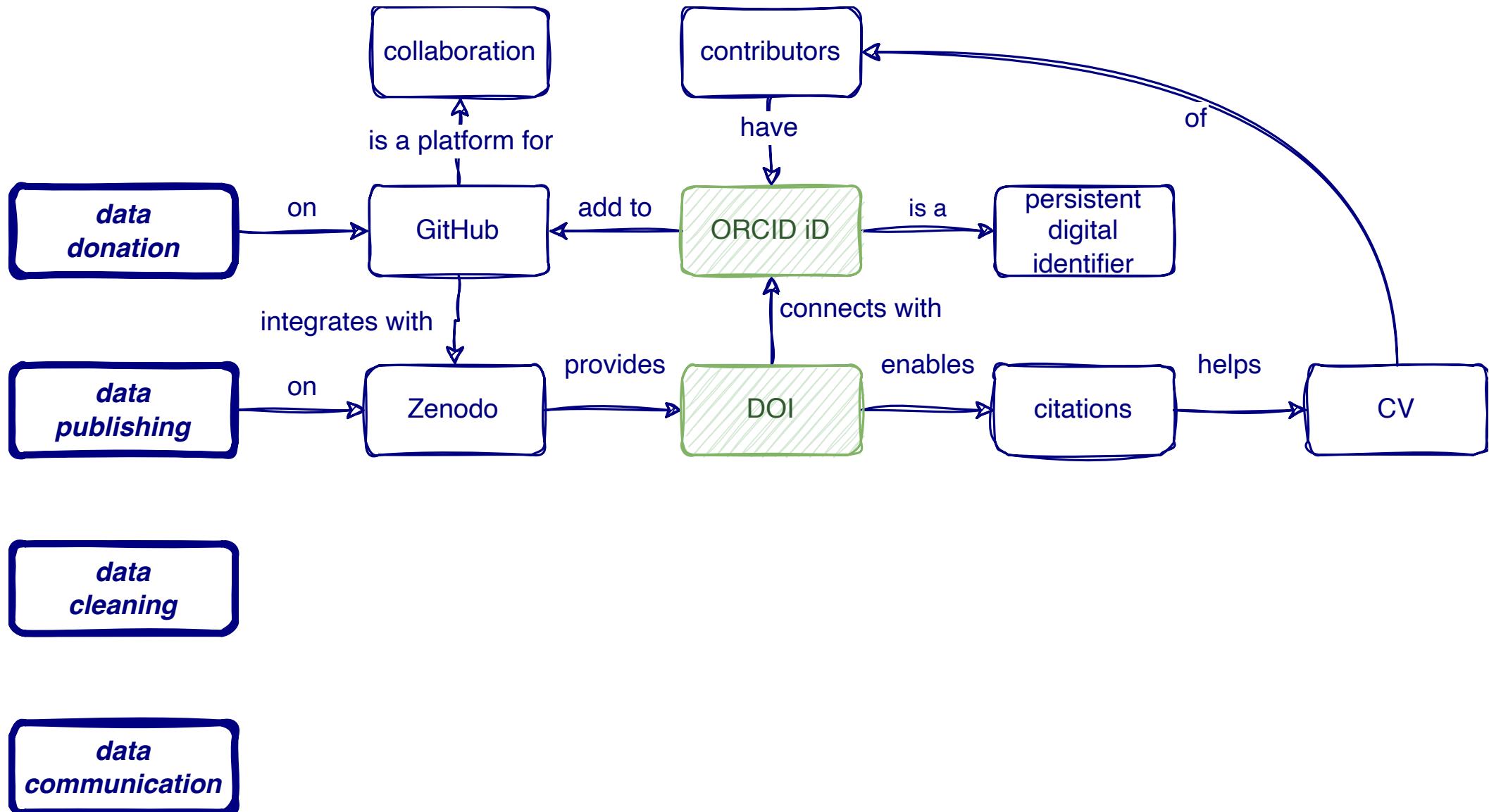
*data
communication*

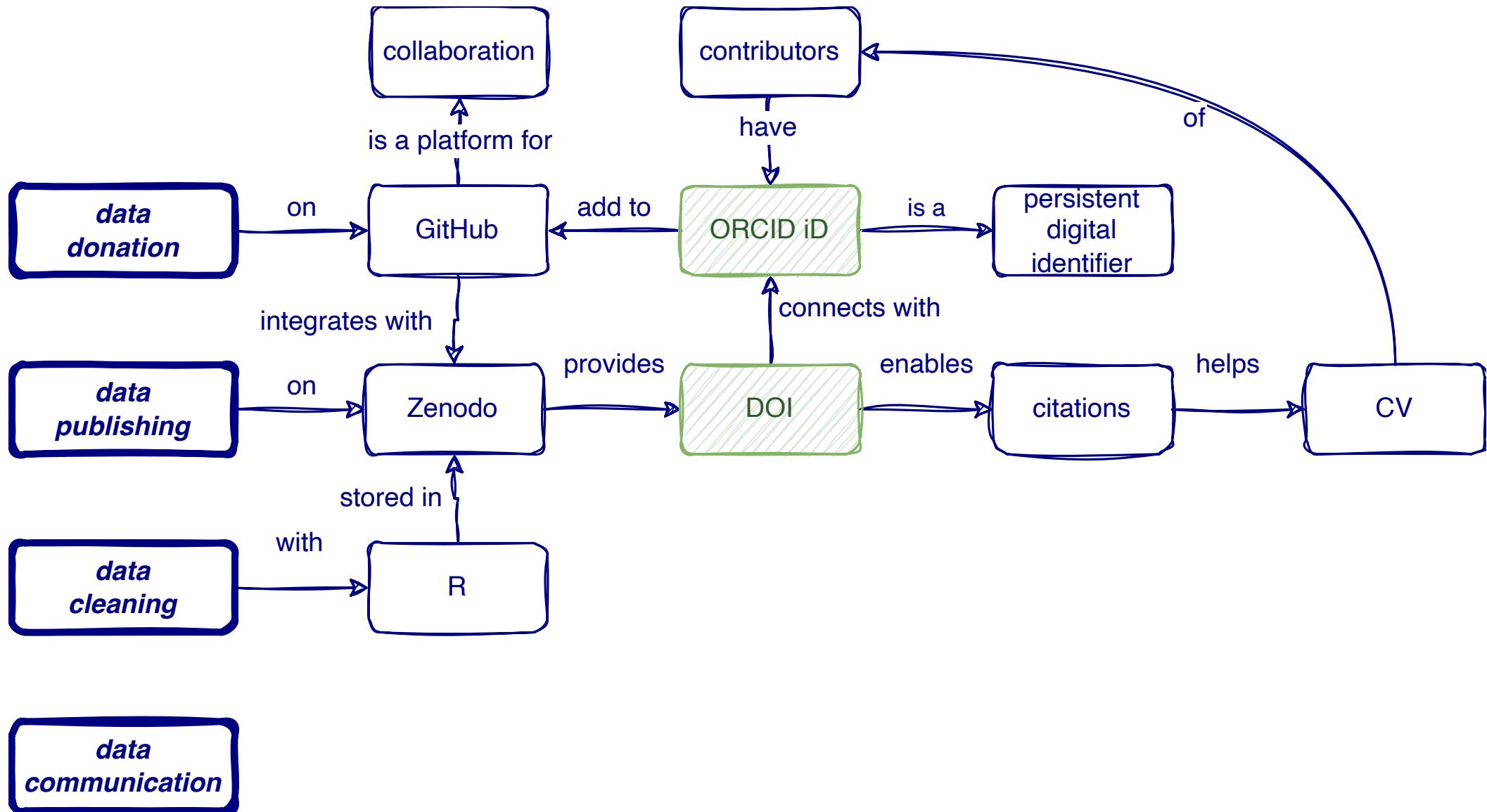


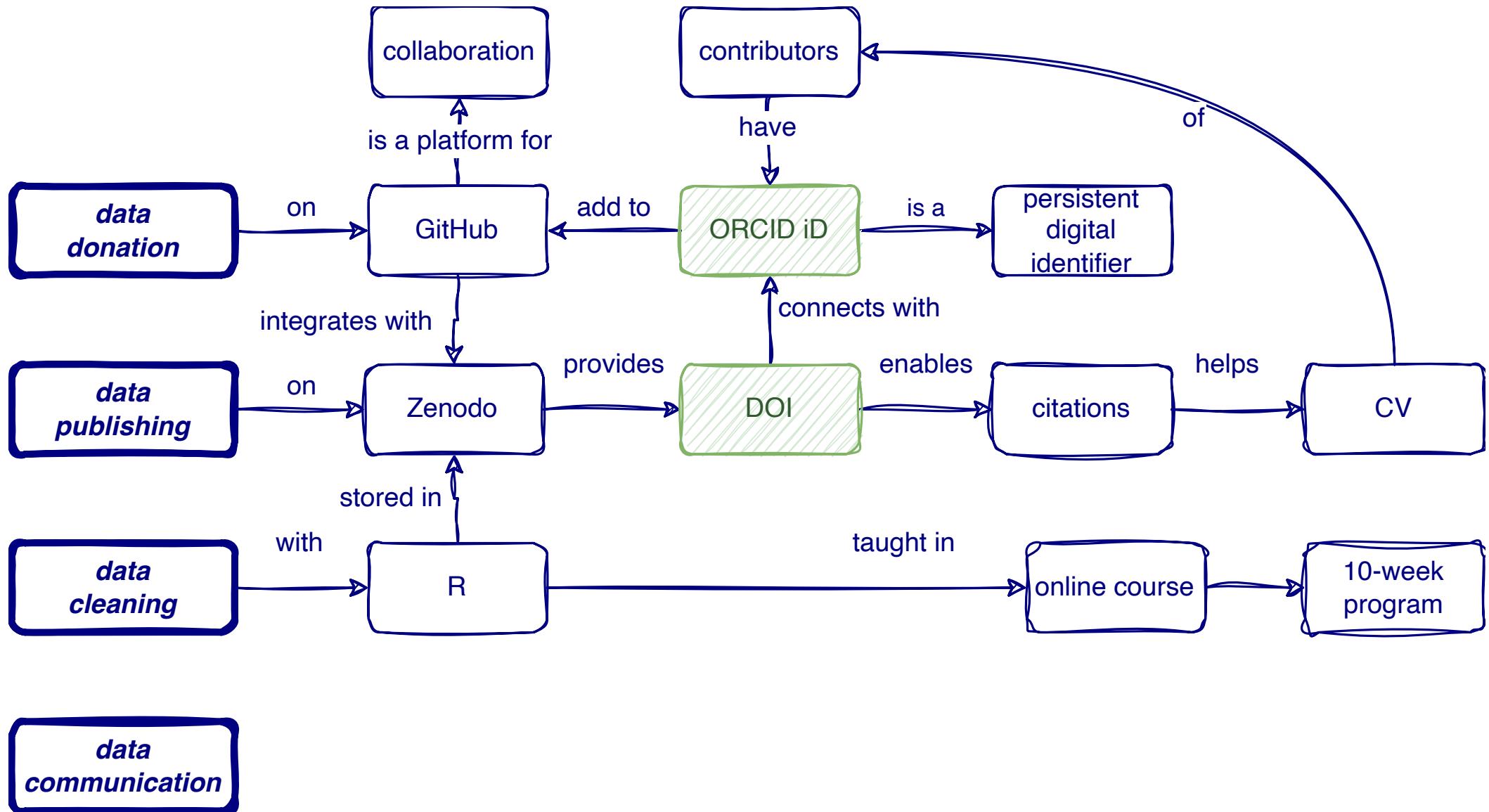


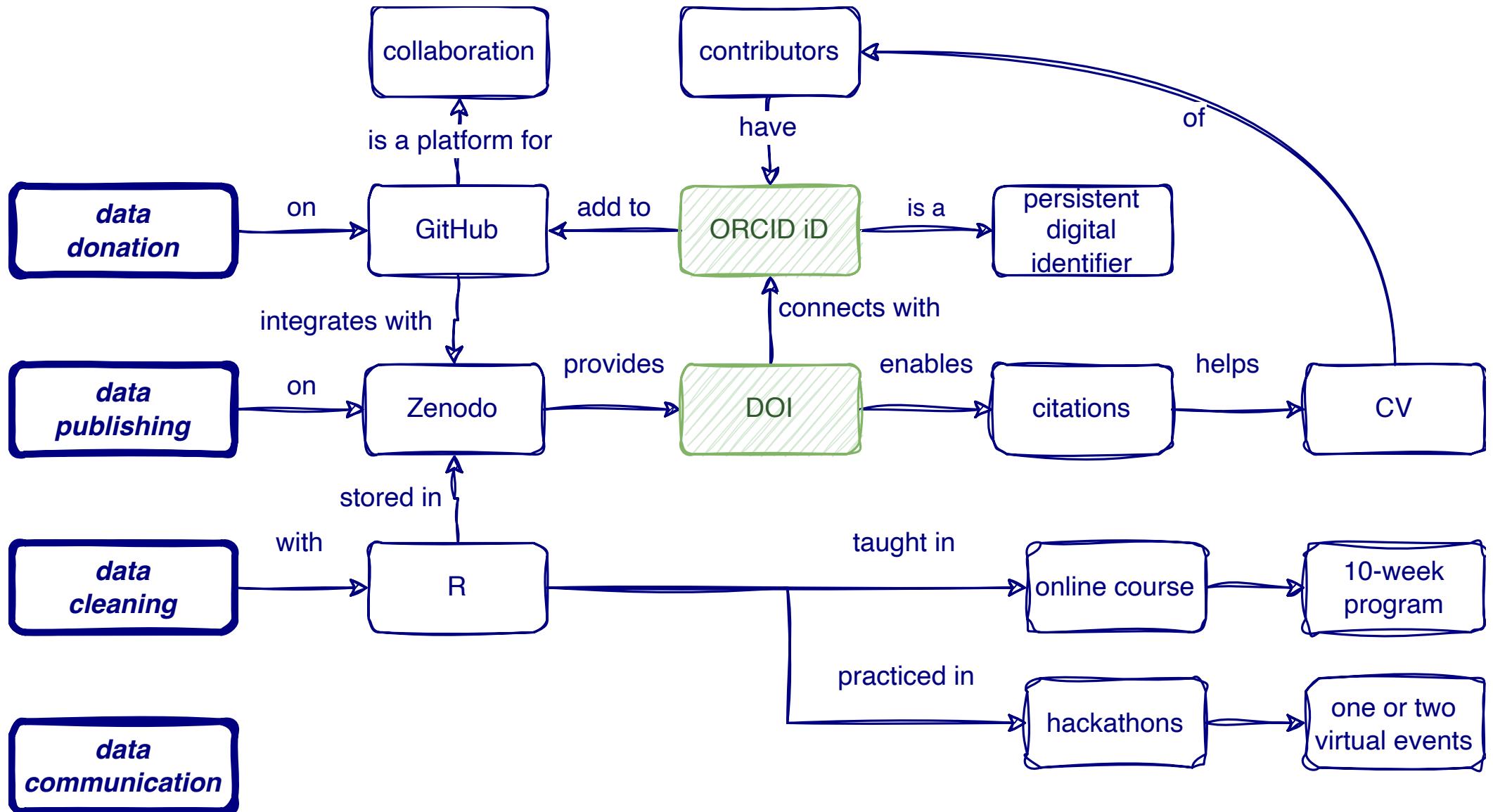


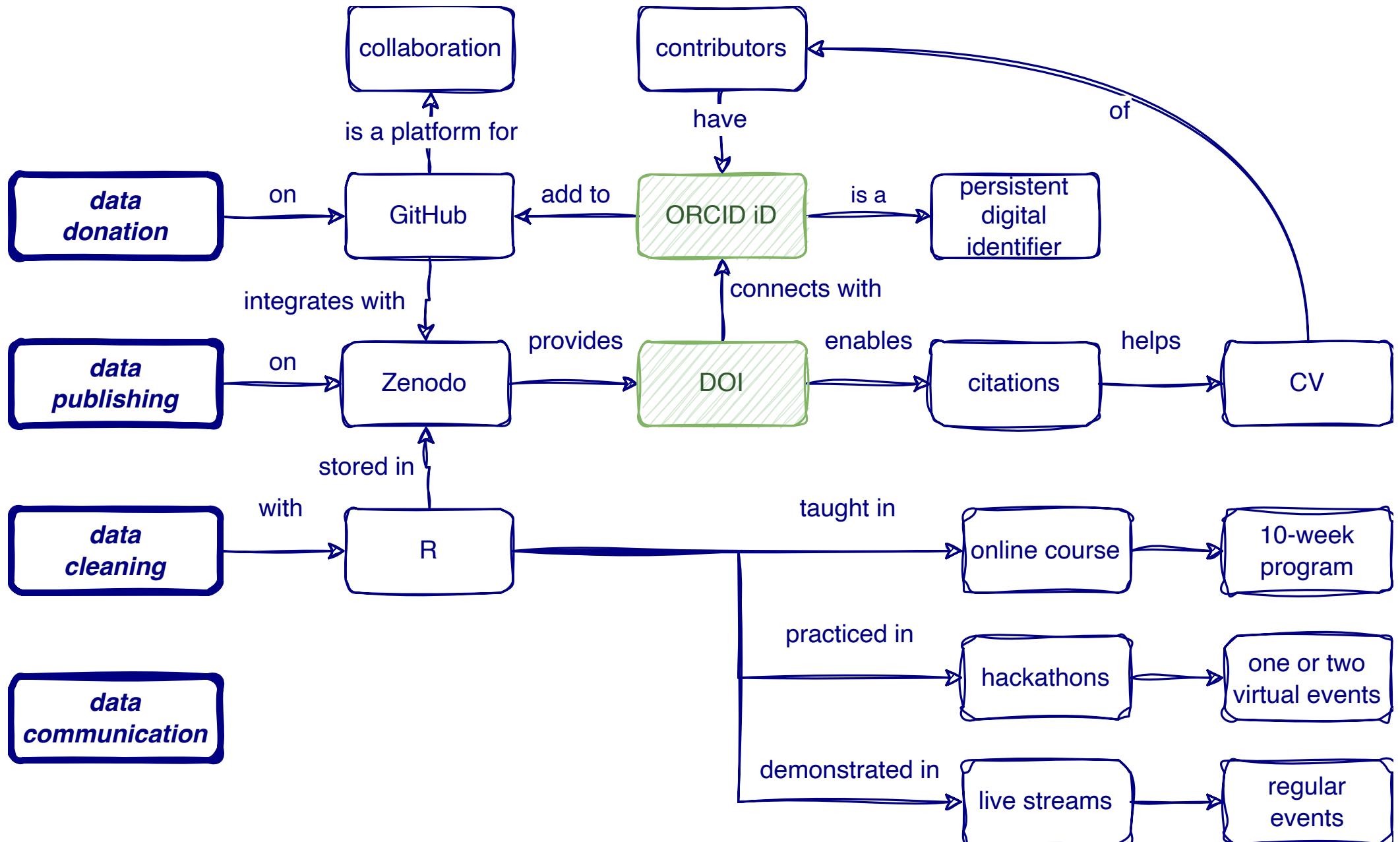


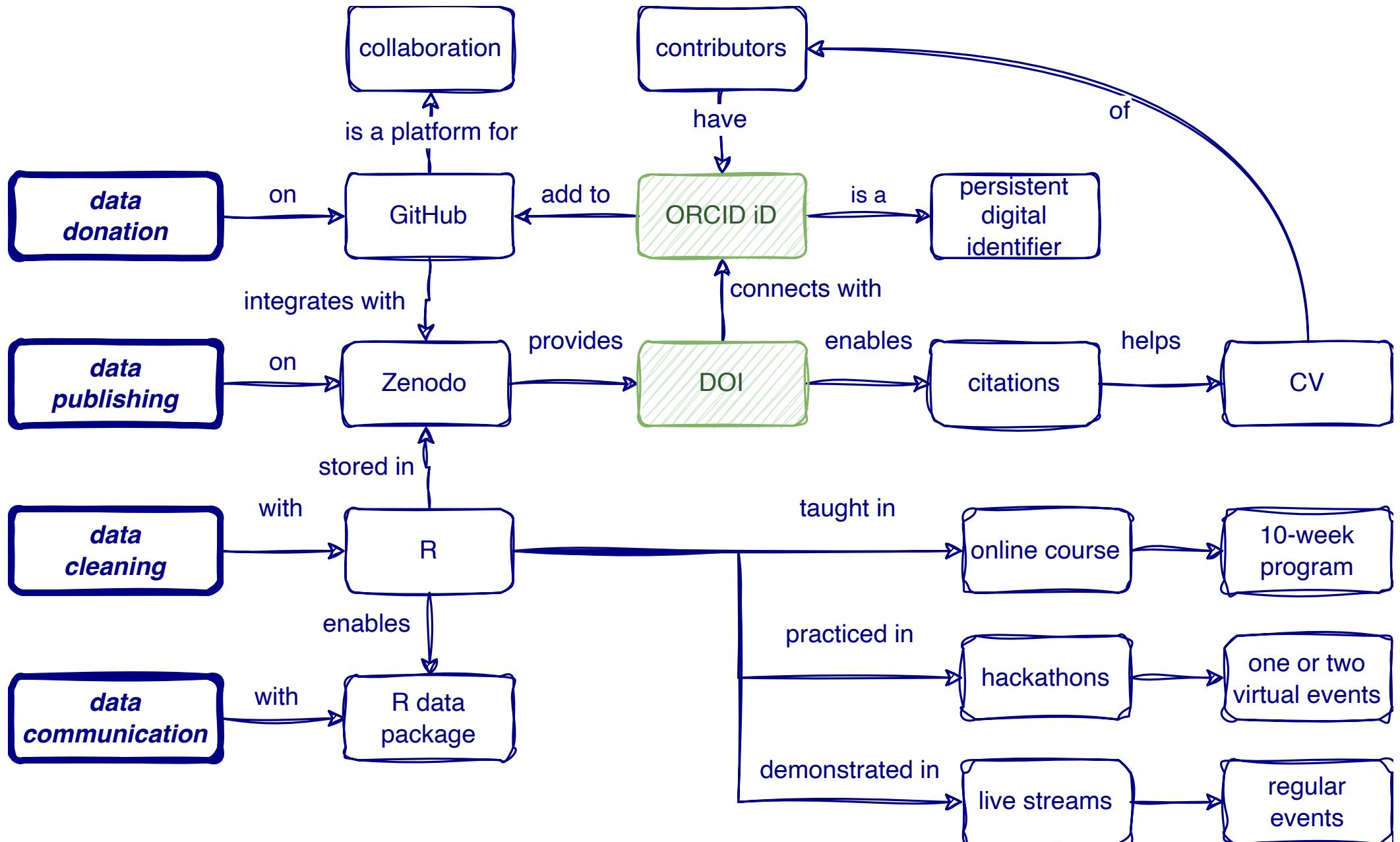


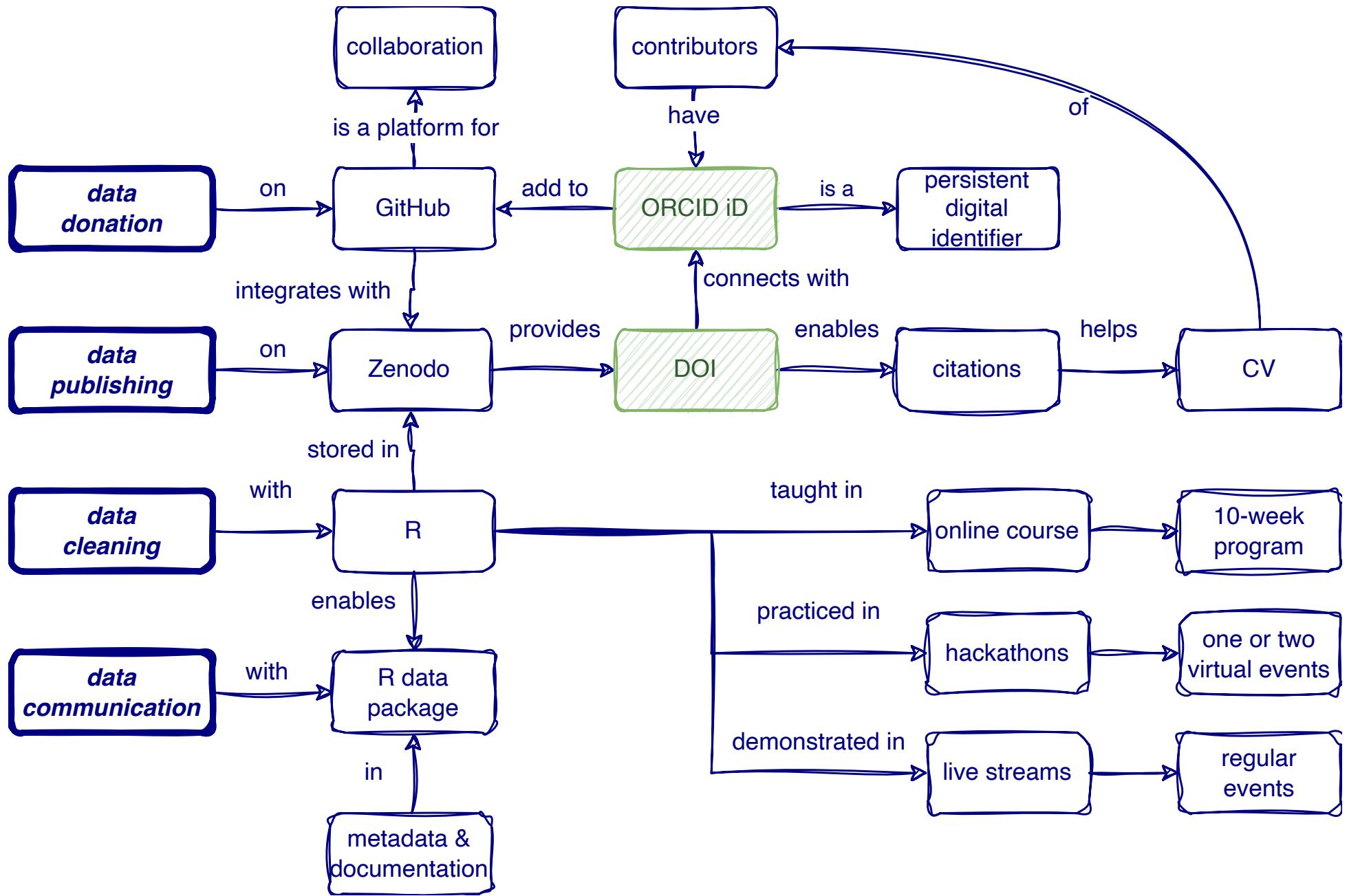


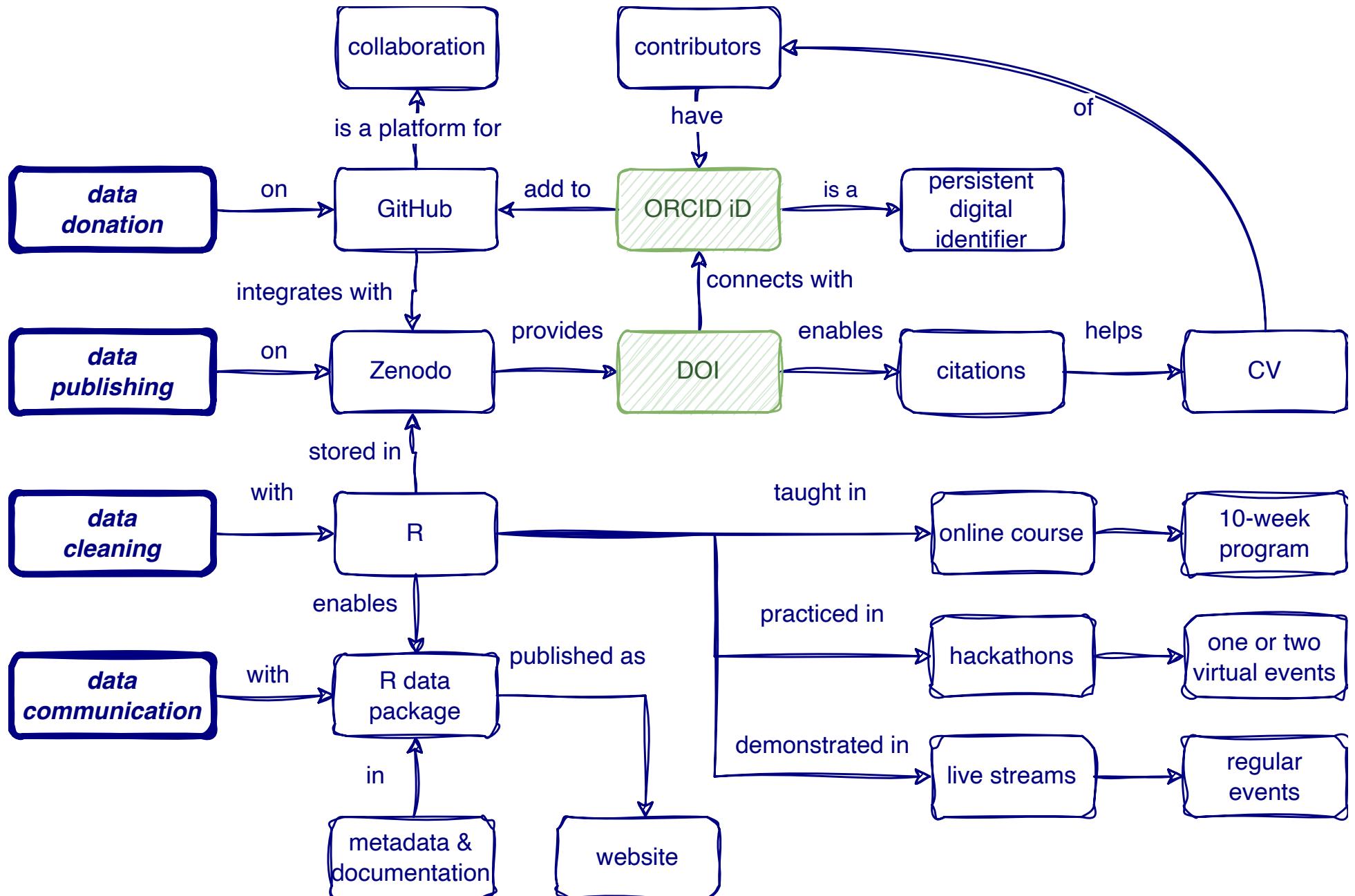


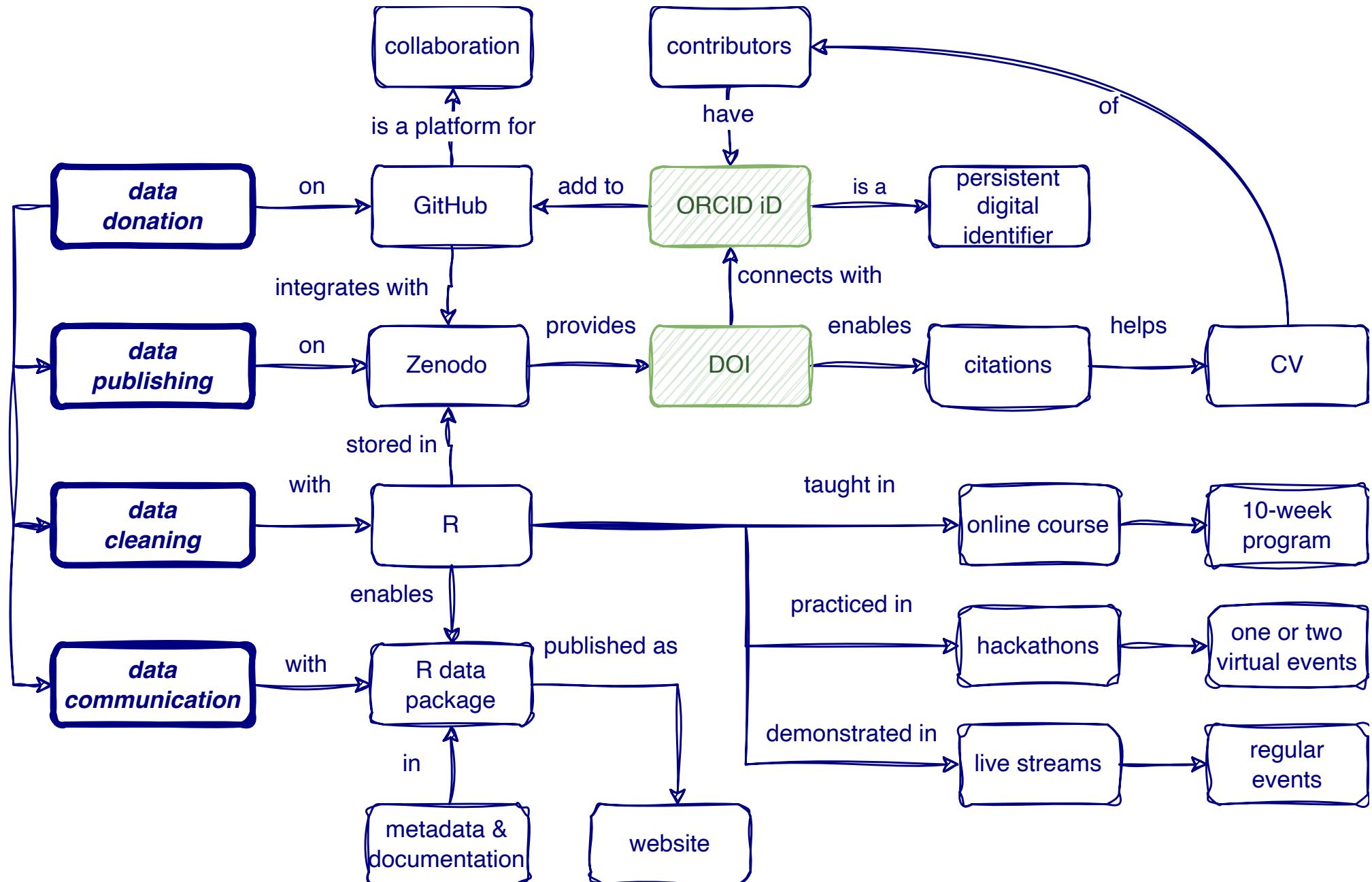












The Product

larnsce.github.io/wash-canada/



What does final look like?

The screenshot shows a web browser window with the title "Durban (South Africa) Plastic Waste" and the URL <https://global-health-engineering.github.io/durbanplasticwaste>. The page content is as follows:

durbanplasticwaste 0.1.0 Reference Articles ▾

durbanplasticwaste

Overview

This packages combines data collected as part of an MSc. Thesis Project and an MSc. Semester Project conducted in Durban, South Africa. The projects were supported by the Global Health Engineering group at ETH Zurich, Switzerland.

Installation

You can install the development version of durbanplasticwaste from [GitHub](#) with:

```
# install.packages("devtools")
devtools::install_github("Global-Health-Engineering/durbanplasticwaste")
```

Alternatively, you can download the individual data sets as a CSV or XLSX file from the table below.

dataset	CSV	XLSX
litterboom_counts	Download CSV	Download XLSX

Projects

MSc. Thesis Project

Evaluating the potential of Extended Producer Responsibility returns for a small local waste collection company
The project aims to evaluate the potential of Extended Producer Responsibility (EPR) returns for a small local waste collection company in Durban, South Africa.

License
[Full license](#)
CC BY 4.0

Citation
[Citing durbanplasticwaste](#)

Developers
Raúl Bergen
Author

Lars Schöbitz
Maintainer

Chiara Meyer-Piening
Author

Global Health Engineering
Funder

[More about authors...](#)

Dev status
DOI [10.5281/zenodo.7708756](#)

Engage

larnsce.github.io/wash-canada/



Our channels

One-way communication

- Website: openwashdata.org
- Newsletter:
buttondown.email/openwashdata

Two-way engagement

- Instant messaging: [Matrix](#) (e.g. with Element chat client)
- Submit data ideas:
github.com/openwashdata/data/issues
- Social media: probably none

Thanks! 🌻

Thanks

A large proportion of these slides are taken from [Mine Çetinkaya Rundel's “Hello Quarto” presentation](#) & [Thomas Mock's “Quarto for the Curious” presentation](#)

Project openwashdata is supported by the [Open Research Data Program of the ETH Board](#).

The slides were created via revealjs and Quarto:

<https://quarto.org/docs/presentations/revealjs/>

You can [view source code of slides on GitHub](#)

Or you can [download slides in PDF format](#)

This material is licensed under [Creative Commons Attribution Share Alike 4.0 International](#).

References

- Bergen, Raúl, Lars Schöbitz, Chiara Meyer-Piening, Boynton Lin, Elizabeth Tilley, Marc Kalina, Siphewe Rakgabale, First Name Last Name, and Global Health Engineering. 2023. “durbanplasticwaste: Durban (South Africa) Plastic Waste Data.” Zenodo. <https://doi.org/10.5281/zenodo.7708756>.
- Greene, Nicola, Sarah Hennessy, Tate W. Rogers, Jocelyn Tsai, and Francis L. de los Reyes III. 2021. “The Role of Emptying Services in Provision of Safely Managed Sanitation: A Classification and Quantification of the Needs of LMICs.” *Journal of Environmental Management* 290 (July): 112612. <https://doi.org/10.1016/j.jenvman.2021.112612>.

larnsce.github.io/wash-canada/

