openwashdata

a community effort to bring open data practices to the WASH sector

Lars Schöbitz

lschoebitz@ethz.ch

Global Health Engineering

January 22, 2024

https://openwashdata.org/pages/gallery/slides/

openwashdata community

https://openwashdata.org/pages/gallery/slides/

openwashdata community

Vision

An active global community that applies FAIR principles (Wilkinson et al. 2016) to data generated in the great water, sanitation, and hygiene sector.

Mission

Empower WASH professionals to engage with tools and workflows for open data and code.

The Opportunity

https://openwashdata.org/pages/gallery/slides/

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

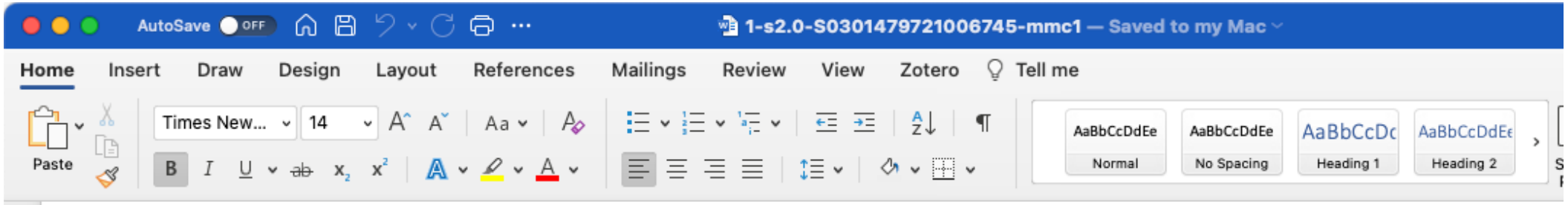


Table 1. The number of onsite sanitation facilities alphabetized by country per Service Type. The Service Type Density is calculated by dividing the population using country population. Thus, the Service Type Density covers users of sewers, onsite facilities, and open defecators. The number of facilities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

	Mechanized				Non-Mechanized				Unemptiable				
Country	No. of Facilities	Service Type Density	Urban	Rural	No. of Facilities	Service Type Density	Urban	Rural	No. of Facilities	Service Type Density	Urban	Rural	No. of Facilities Required
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

PDF reports

https://openwashdata.org/pages/gallery/slides/

PDF reports

Table 2. Influent and effluent qualities of wastewater treated at Duri Kosambi FSTP plant in 2019, as compared to effluent standards

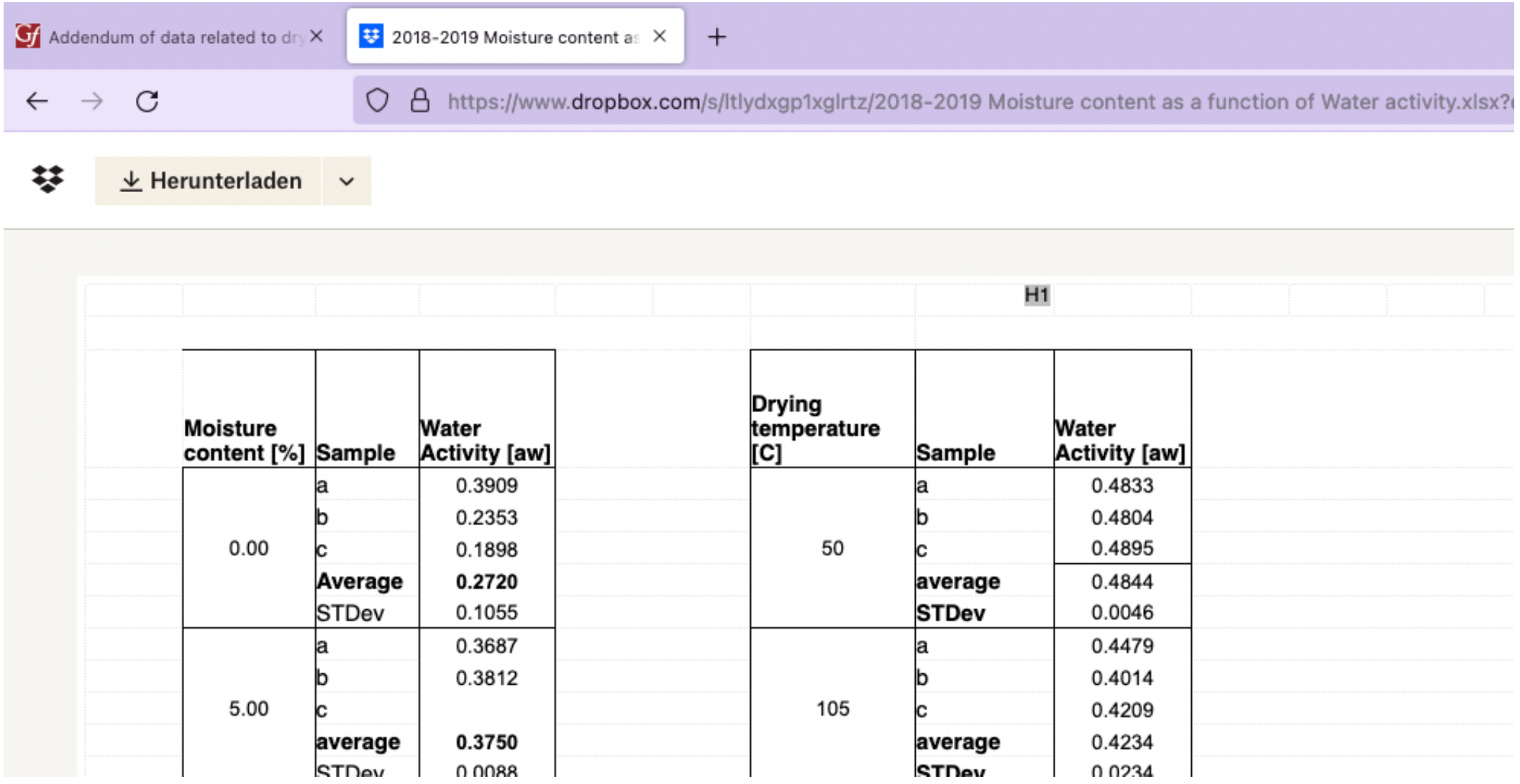
Parameter	Inlet	Outlet
pH	6, 45-7, 88 pH	7, 12-7, 61 pH
Total suspended solids, TSS	340-8933, 33 mg/L	22, 5-84, 29 mg/L
Biochemical oxygen demand, BOD_5	106, 38-646, 82 mg/L	2, 76-69, 79 mg/L
Chemical oxygen demand, COD	687, 9-2780, 37 mg/L	41, 25-127, 67 mg/L
Total organic matter, $KMnO_4$	108, 04-568, 72 mg/L	54, 21-150, 50 mg/L
Ammonia, NH3-N	108, 75-239, 25 mg/L	0, 45-29, 81 mg/L
Methylene blue active surfactant, MBAS	0, 74-2, 69 mg/L	0, 13-0, 78 mg/L

PDF reports + Dropbox

Physiochemical properties ***Addendum of data***

General information	
Type of data	Composition
Place of experimentation	Pollution Research Group, University of KwaZulu-Natal (South Africa)
Dates of the experiments	2018-2019
Feedstock	
Type of faecal material	Faecal sludge from anaerobic baffled reactor (ABR) from a decentralised wastewater treatment plant (DEWAT)
Location of collection	Durban, South Africa
Age before collection	Unknown
Moisture content	~ 90%wt

PDF reports + Dropbox



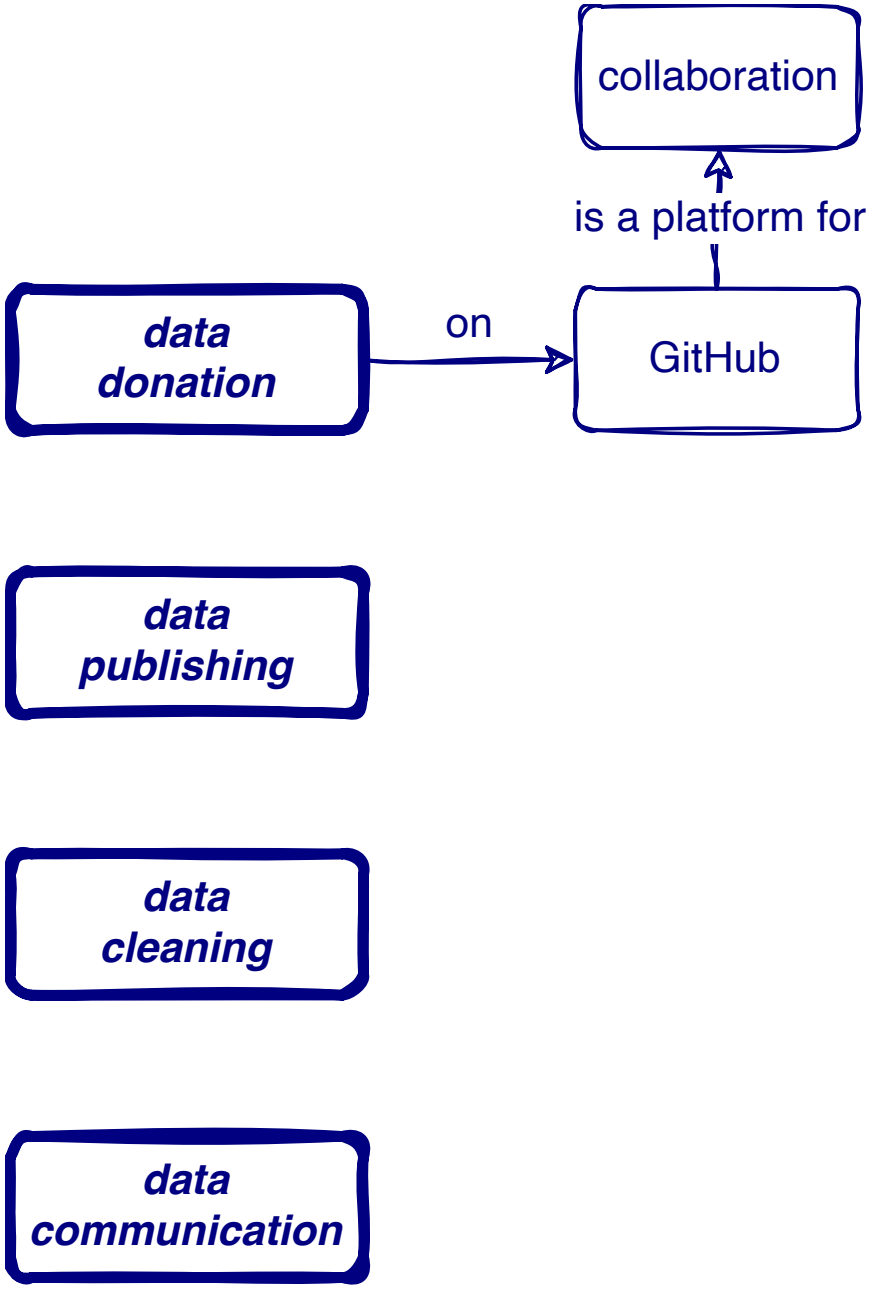
Moisture content [%]	Sample	Water Activity [aw]
0.00	a	0.3909
	b	0.2353
	c	0.1898
	Average	**0.2720**
	STDev	0.1055
5.00	a	0.3687
	b	0.3812
	c	
	average	**0.3750**
	STDev	0.0088

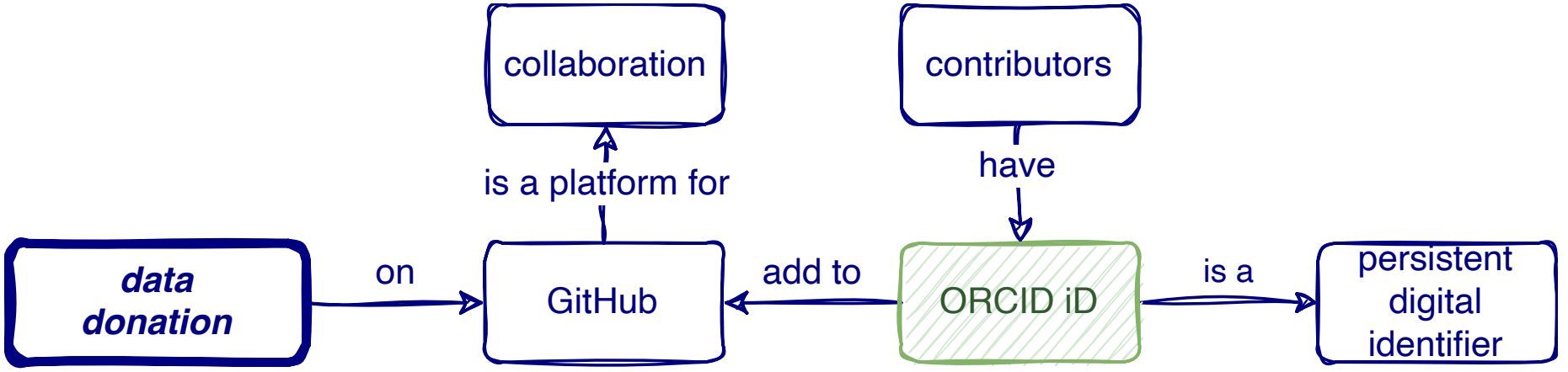
Drying temperature [C]	Sample	Water Activity [aw]
50	a	0.4833
	b	0.4804
	c	0.4895
	average	0.4844
	STDev	0.0046
105	a	0.4479
	b	0.4014
	c	0.4209
	average	0.4234
	STDev	0.0234

The Journey

https://openwashdata.org/pages/gallery/slides/


data donation
data publishing
data cleaning
data communication

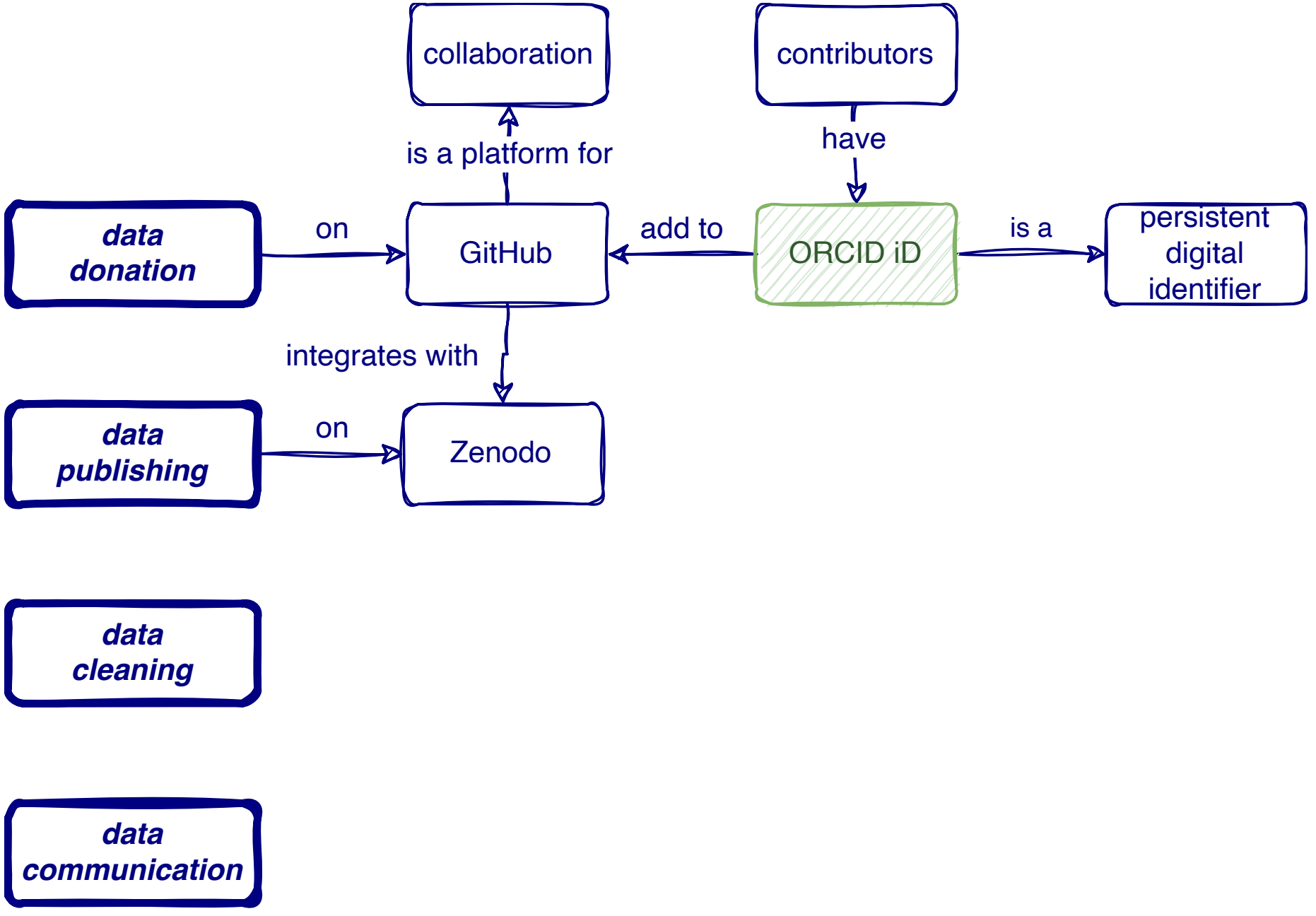

collaboration
is a platform for
data donation
on
GitHub
data publishing
data cleaning
data communication

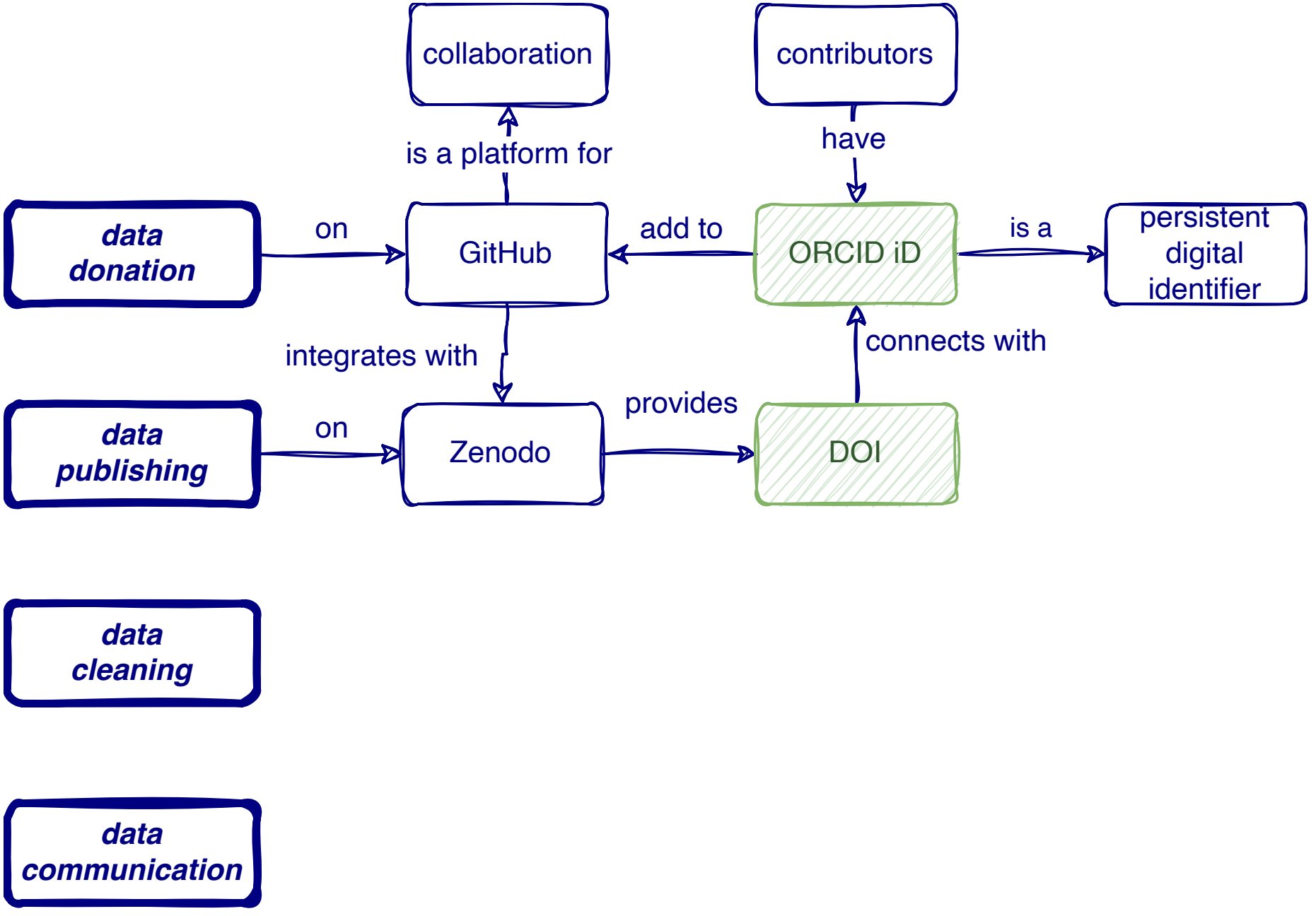


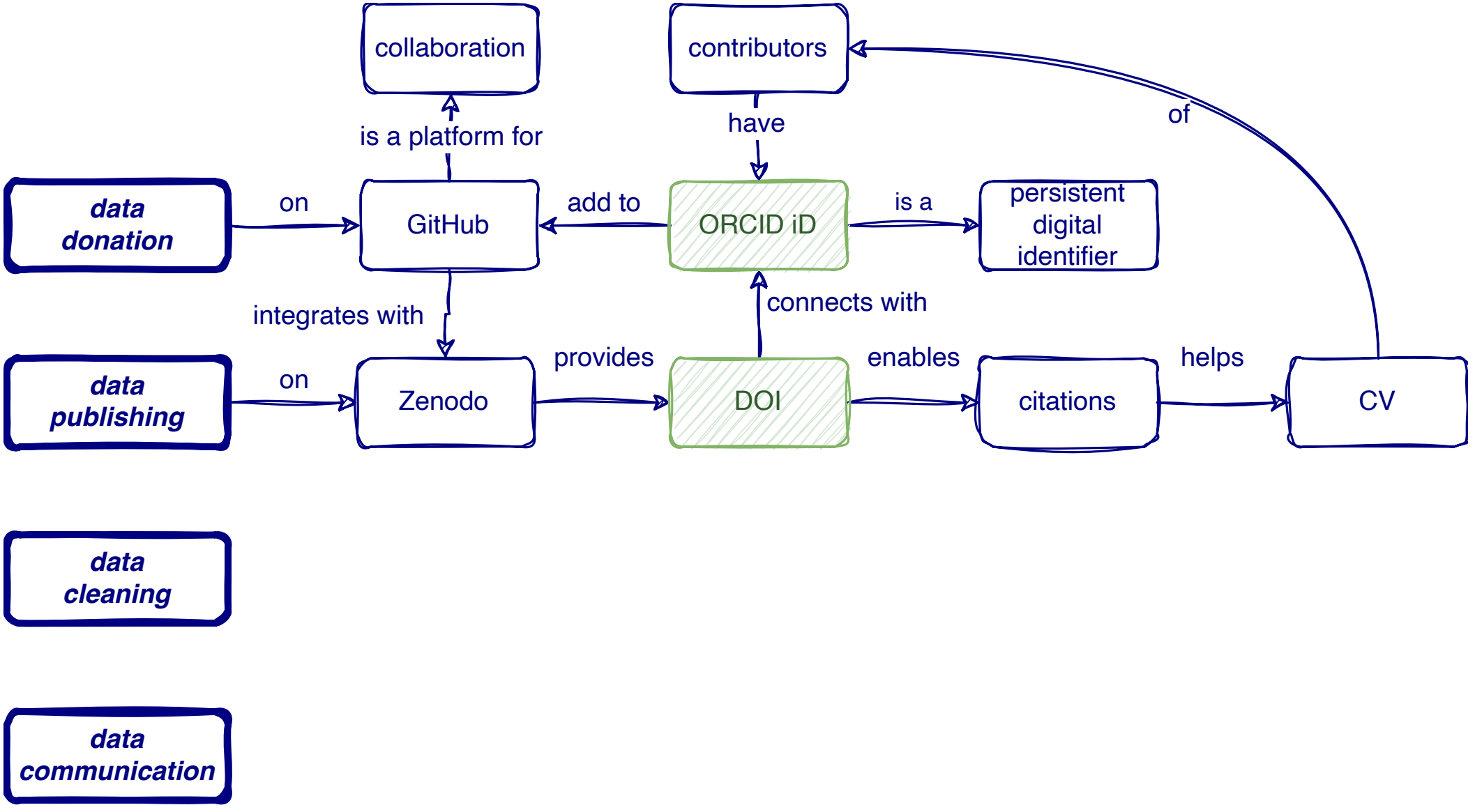
data publishing

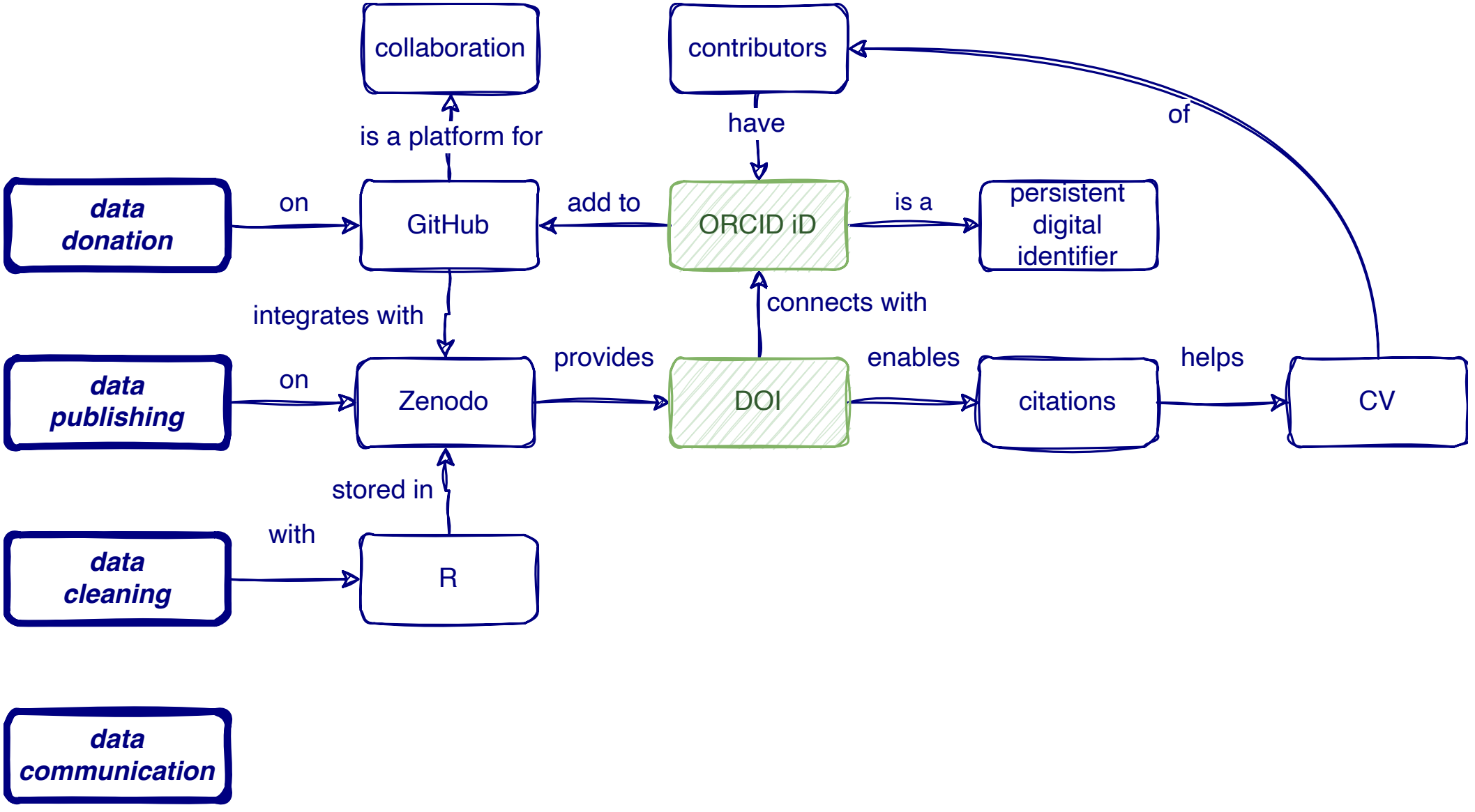
data cleaning

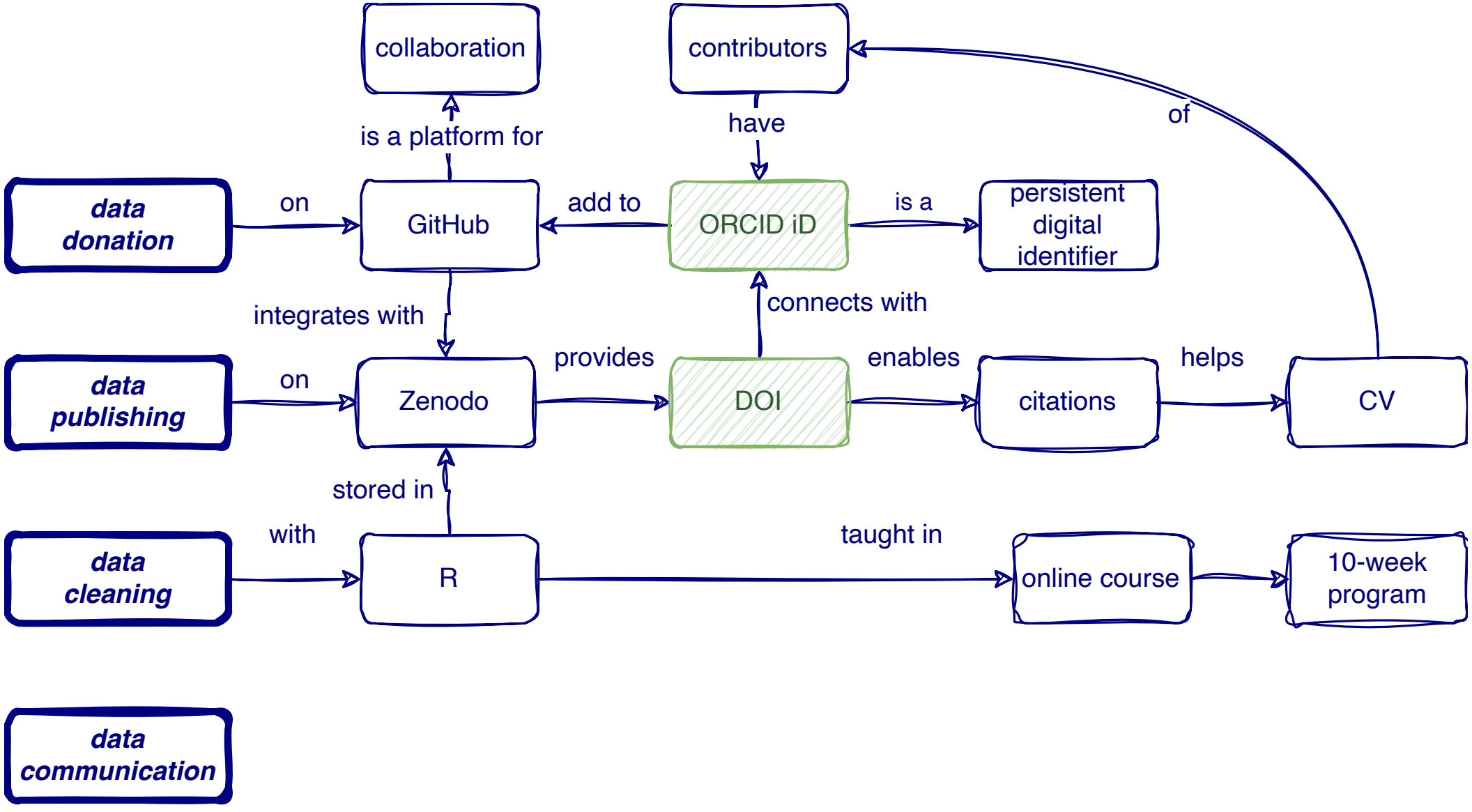
data communication

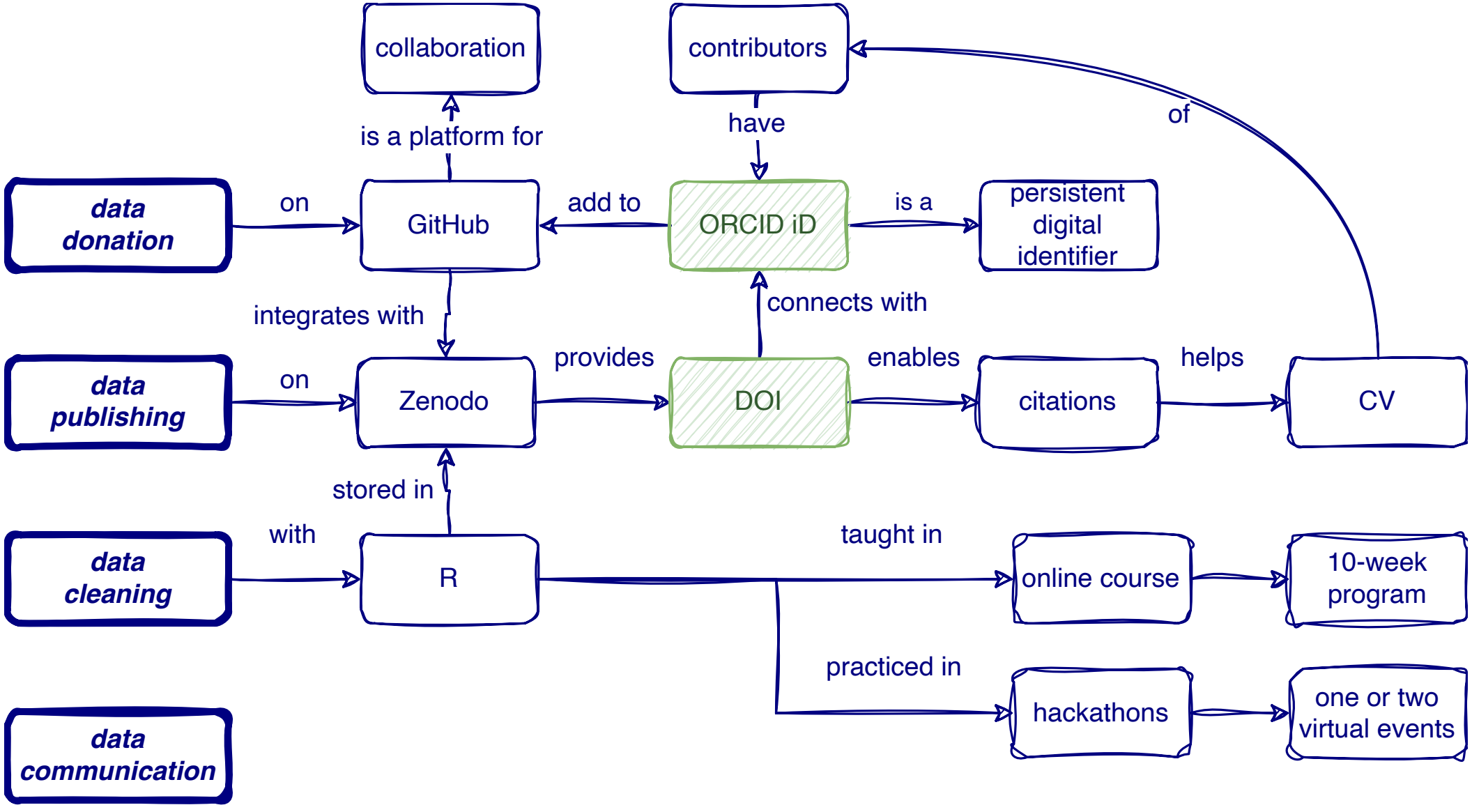

collaboration
contributors
is a platform for
have
data donation
on
GitHub
add to
ORCID iD
is a
persistent digital identifier
integrates with
data publishing
on
Zenodo
data cleaning
data communication

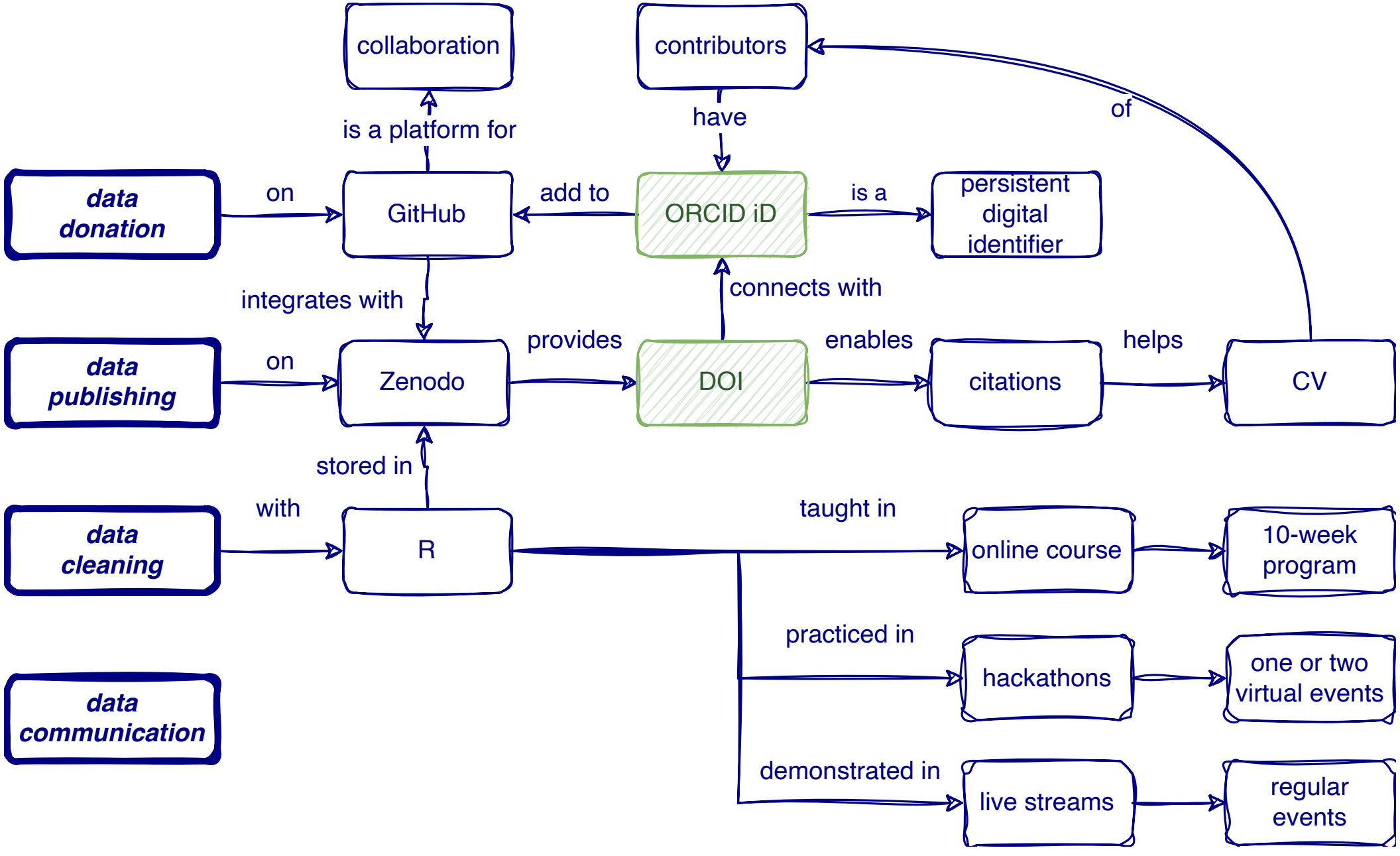

collaboration
contributors
is a platform for
have
data donation
on
GitHub
add to
ORCID iD
is a
persistent digital identifier
integrates with
connects with
data publishing
on
Zenodo
provides
DOI
data cleaning
data communication

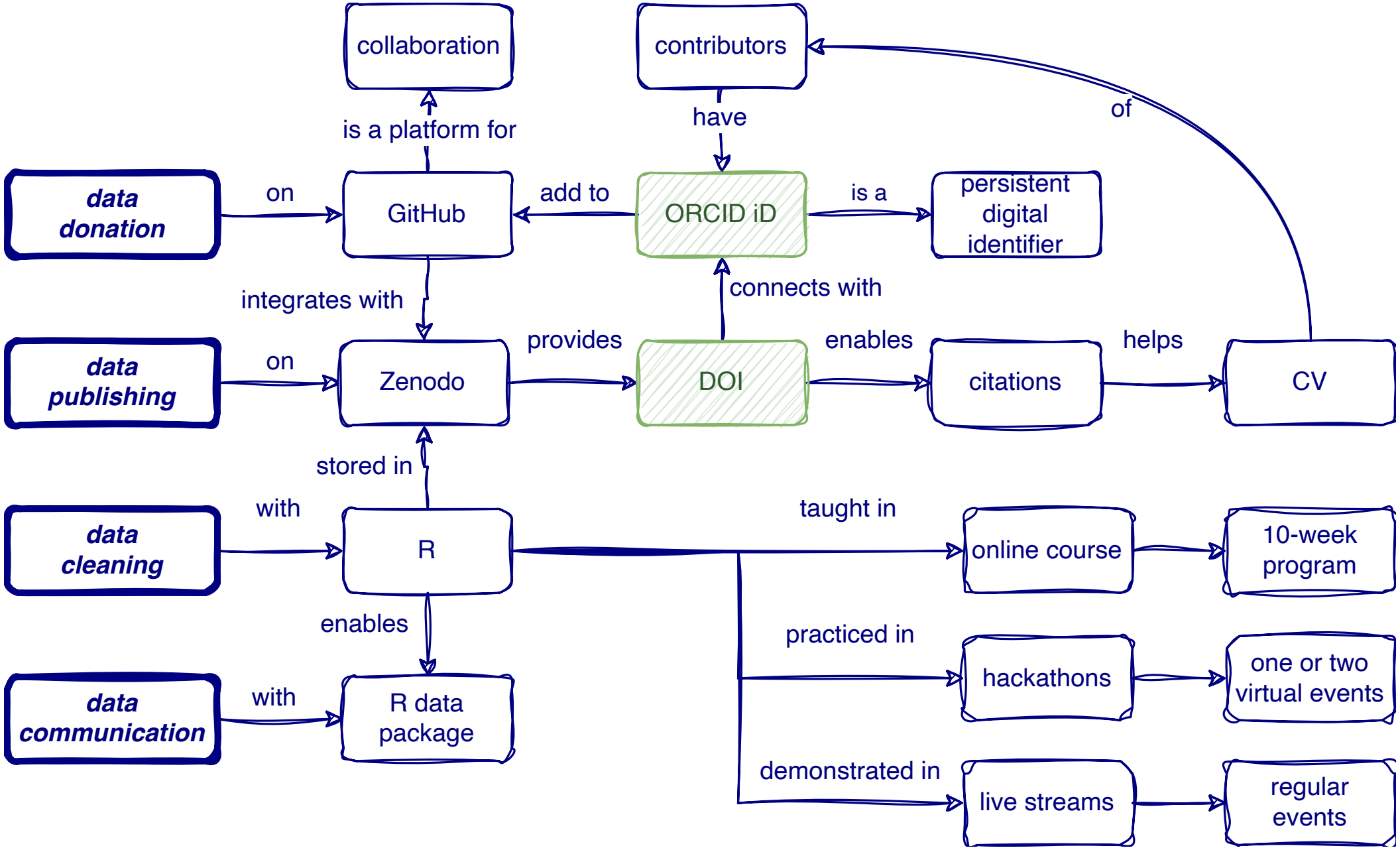

collaboration
contributors
is a platform for
have
of
data donation
on
GitHub
add to
ORCID iD
is a
persistent digital identifier
integrates with
connects with
data publishing
on
Zenodo
provides
DOI
enables
citations
helps
CV
data cleaning
data communication

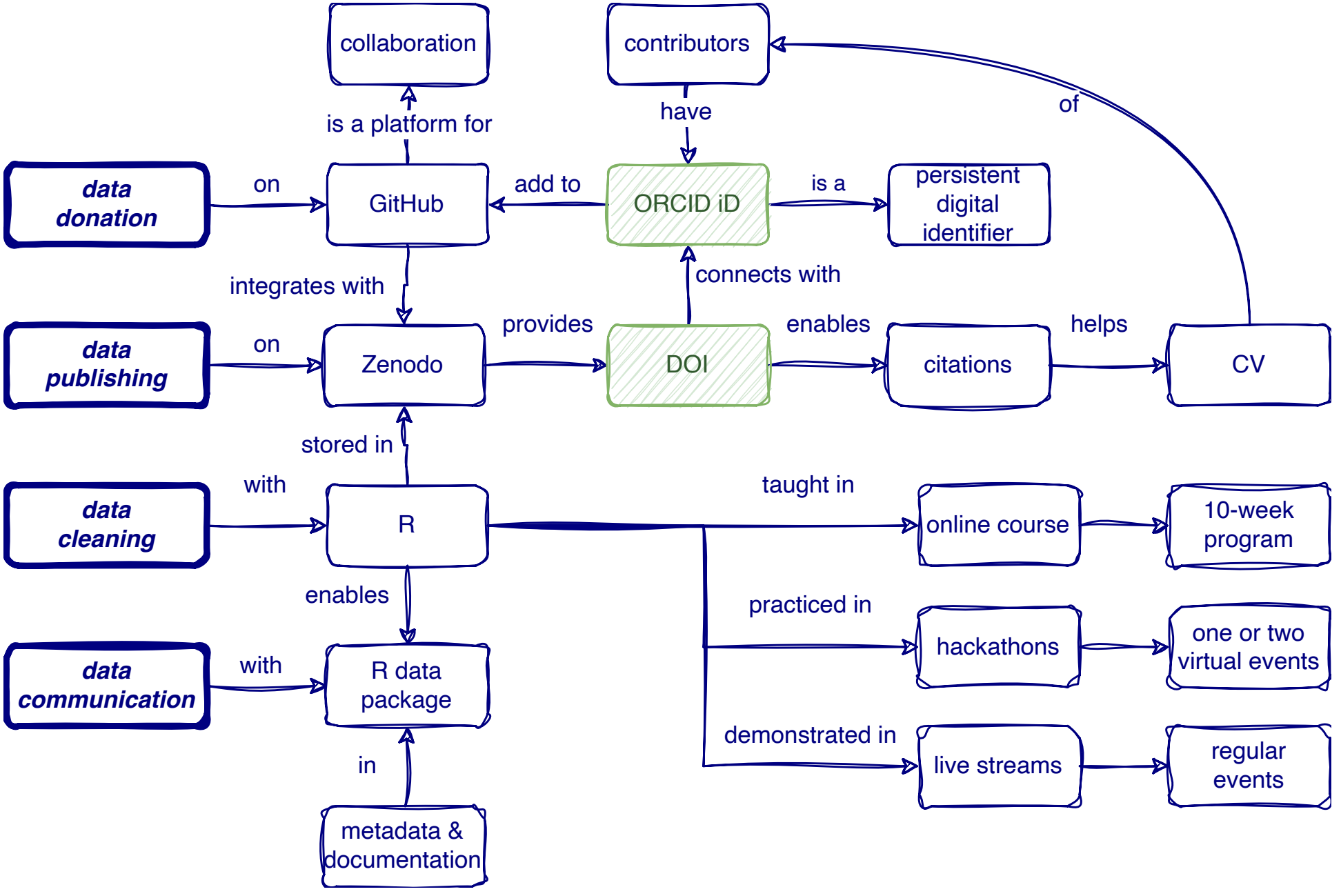

collaboration
contributors
is a platform for
have
of
data donation
on
GitHub
add to
ORCID iD
is a
persistent digital identifier
integrates with
connects with
data publishing
on
Zenodo
provides
DOI
enables
citations
helps
CV
stored in
with
data cleaning
R
data communication

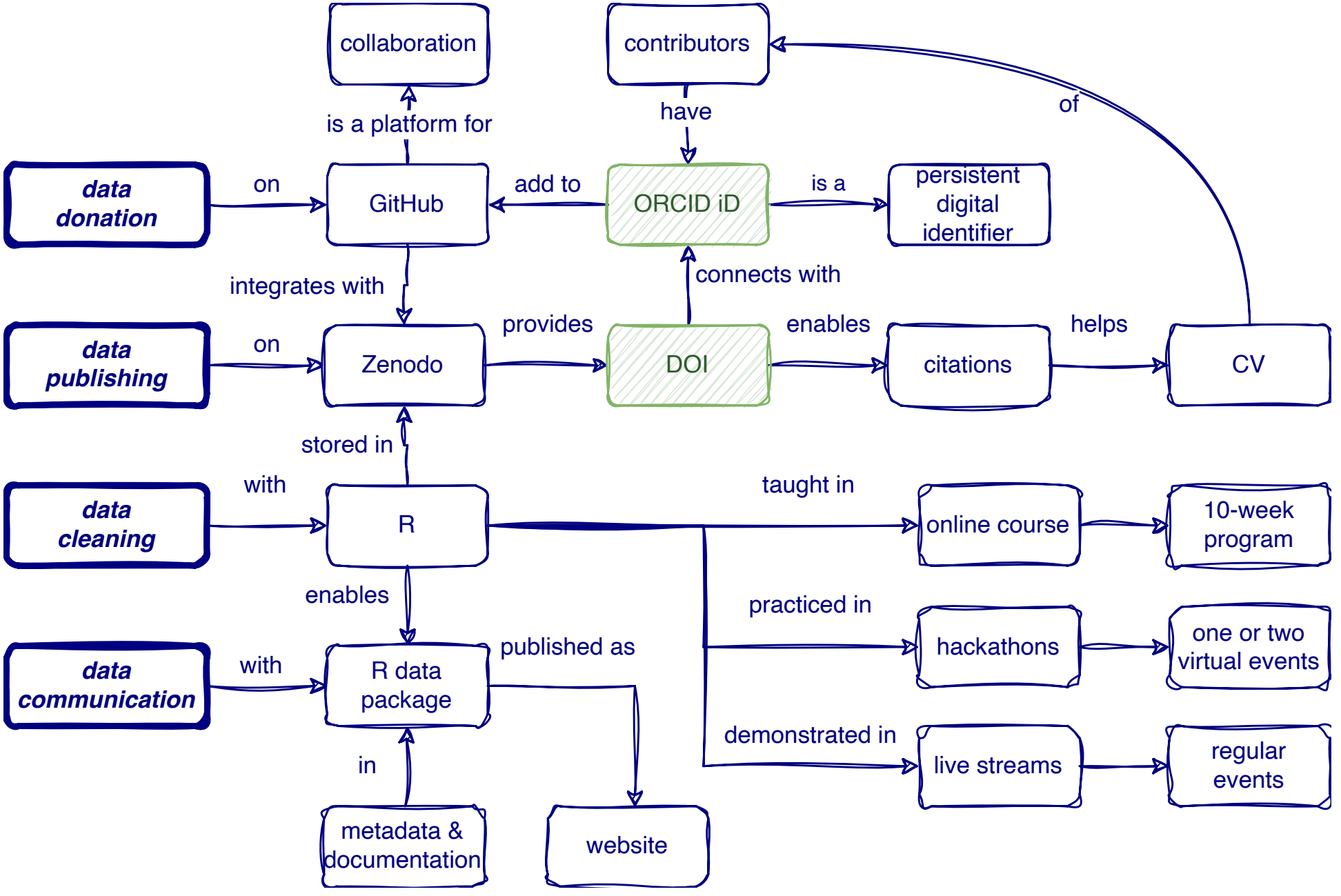

collaboration
contributors
is a platform for
have
of
data donation
on
GitHub
add to
ORCID iD
is a
persistent digital identifier
integrates with
connects with
data publishing
on
Zenodo
provides
DOI
enables
citations
helps
CV
stored in
data cleaning
with
R
taught in
online course
10-week program
data communication


collaboration
contributors
is a platform for
have
of
data donation
on
GitHub
add to
ORCID iD
is a
persistent digital identifier
integrates with
connects with
data publishing
on
Zenodo
provides
DOI
enables
citations
helps
CV
stored in
with
data cleaning
R
taught in
online course
10-week program
practiced in
hackathons
one or two virtual events
data communication

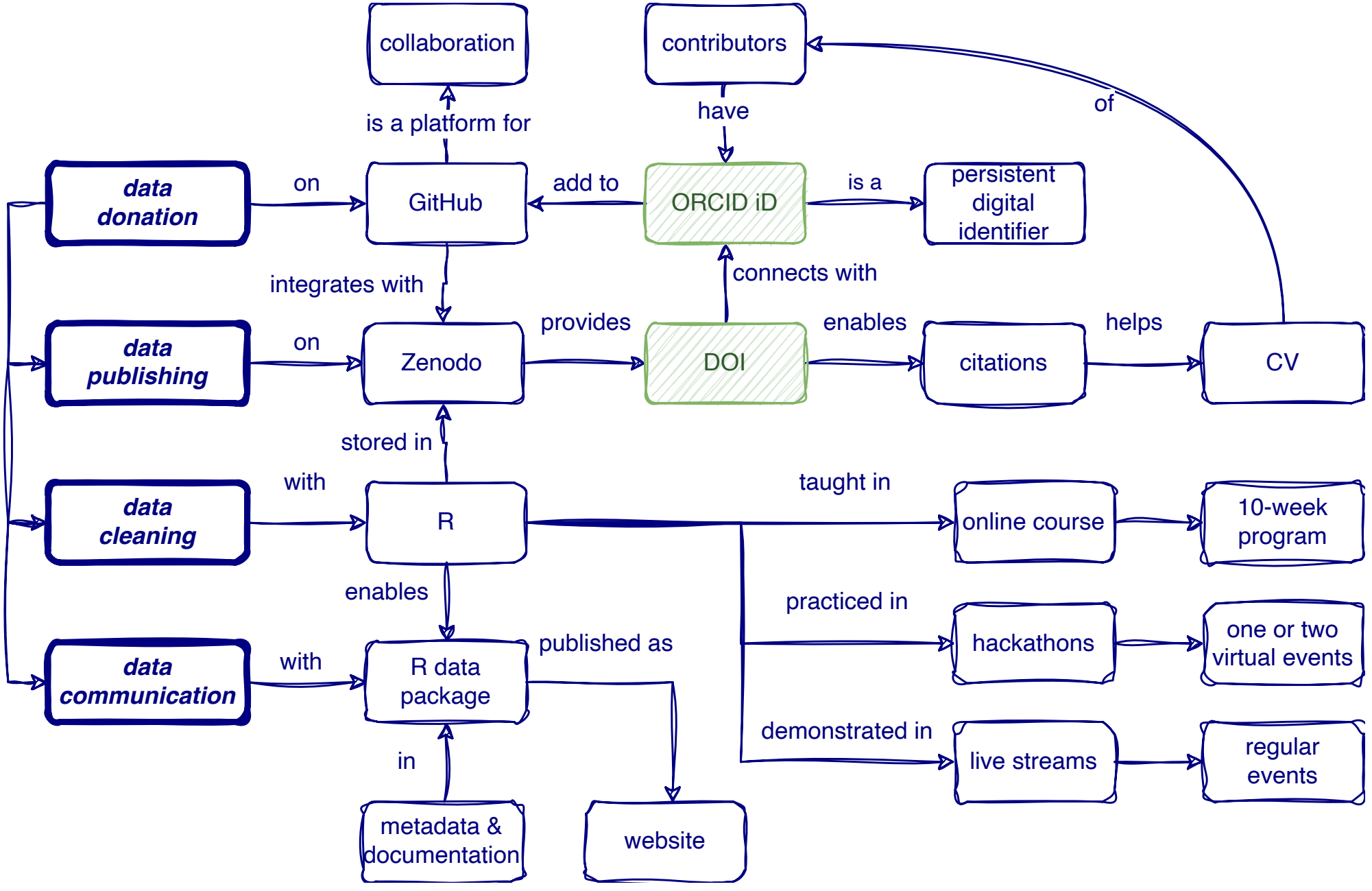

collaboration
contributors
is a platform for
have
of
data donation
on
GitHub
add to
ORCID iD
is a
persistent digital identifier
integrates with
connects with
data publishing
on
Zenodo
provides
DOI
enables
citations
helps
CV
stored in
data cleaning
with
R
taught in
online course
10-week program
practiced in
hackathons
one or two virtual events
data communication
demonstrated in
live streams
regular events


collaboration
contributors
is a platform for
have
of
data donation
on
GitHub
add to
ORCID iD
is a
persistent digital identifier
integrates with
connects with
data publishing
on
Zenodo
provides
DOI
enables
citations
helps
CV
stored in
data cleaning
with
R
taught in
online course
10-week program
enables
practiced in
hackathons
one or two virtual events
data communication
with
R data package
demonstrated in
live streams
regular events


collaboration
contributors
is a platform for
have
of
data donation
on
GitHub
add to
ORCID iD
is a
persistent digital identifier
integrates with
connects with
data publishing
on
Zenodo
provides
DOI
enables
citations
helps
CV
stored in
data cleaning
with
R
taught in
online course
10-week program
enables
practiced in
data communication
with
R data package
hackathons
one or two virtual events
demonstrated in
in
live streams
regular events
metadata & documentation



https://openwashdata.org/pages/gallery/slides/



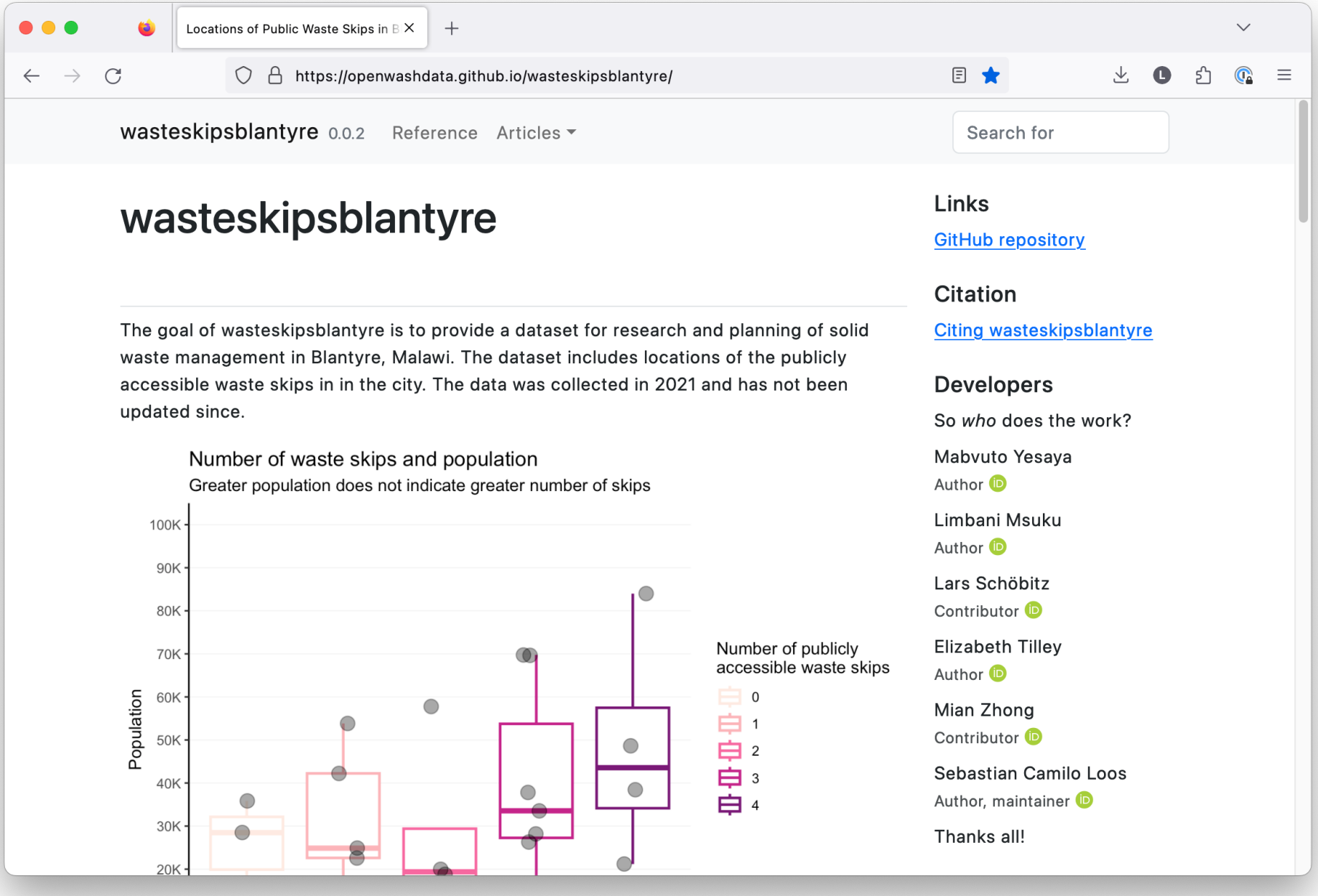
https://openwashdata.org/pages/gallery/slides/

The Product

https://openwashdata.org/pages/gallery/slides/

What does final look like?

https://openwashdata.org/pages/gallery/slides/



https://openwashdata.org/pages/gallery/slides/

Engage

https://openwashdata.org/pages/gallery/slides/

Our channels

One-way communication

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

Two-way engagement

- Instant messaging: Element based on Matrix Chat | openwashdata-lobby | ghe-open
- Data donation ideas: github.com/openwashdata/data/issues
- Social media: Global Health Engineering LinkedIn

course: data science for openwashdata

ds4owd-001.github.io/website/

Zoom for 10 modules over 2 months at the following times:

- Start: 31st October 2023 - 2 pm to 4:30 pm CET
- End: 20th February 2024 - 2 pm to 4:30 pm CET

Registration open for next course: https://forms.gle/AhhWpPfnbLwzp5Ai9

- free
- provides participants with a certificate
- using exclusively tools that are free and open source
- offers 1:1 coding support for a final project with own data

course: data science for openwashdata 001

- 200 registrations
- 110 show-ups
- expected 40 graduates all with a reproducible data analysis report (paper)
- motivating graduates to publish underlying data with openwashdata

open - misconceptions

open - misconceptions

- Misconception 1: Publishing my data does not benefit anyone
- Misconception 2: Others may criticize my code
- Misconception 3: Publishing my content under CC-BY-NC will prevent people from exploiting my content commercially. (go for CC-BY)

> The only way to write good code is to write tons of shitty code first. Feeling shame about bad code stops you from getting to good code. - Hadley Wickham (Chief Scientist, Posit PBC)

Thanks 🌻

This project was 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

References

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.

Mupinga, Ratidzaishe T, Tanaka M Chatema, Savanna R Perumal, Eva Mary, et al. 2021. “Addendum of Data Related to Drying of Faecal Sludge from on-Site Sanitation Facilities and Fresh Faeces.” *Gates Open Res* 4 (188): 188.

Soeters, S, P Mukheibir, and J Willetts. 2021. “Treatment Technologies in Practice: On-the-Ground Experiences of Faecal Sludge and Wastewater Treatment.”

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. “The FAIR Guiding Principles for Scientific Data Management and Stewardship.” *Scientific Data* 3 (1). https://doi.org/10.1038/sdata.2016.18.

Yesaya, Mabvuto, Limbani Msuku, Elizabeth Tilley, and Sebastian Camilo Loos. 2023. “Wasteskipsblantyre: Locations of Public Waste Skips in Blantyre, Malawi.” https://doi.org/10.5281/zenodo.6470427.