Universiteit Utrecht

SUMMERFAIR
combining data science and infection models for estimating transmission parameters

Egil Fischer & Giske Lagerweij
Faculty of Veterinary Medicine

Transmission in small populations

Photo Dr. Anita Dame-Korevaar

Context:
Transmission experiment with ESBL-producing E. coli.
Challenge at 8:00, swaps taken at 16:00 daily
Samples are cloaca swaps put on plate after standard dilution series
- = no growth on selective plate, 1-5 = log concentration ESBL-producing E. coli, <1 is positive but too little to count, >6 positive but too many to count

		Sample_time (days post challenge)							
Host_id	**Challenge**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	Yes	2	5	>6	4	3	4	-	-
2	Yes	1	-	-	-	4	5	6	5
3	Yes	<1	5	4	2	1	-	<1	-
4	No	-	-	4	3	-	4	4	-
5	No	-	<1	2	-	-	3	-	-
6	No	-	-	4	3	-	-	-	<1

Infection status of an animal?

- Biological / epidemiological factors:
  - Passage of inoculum
  - Intermittent shedding
  - Sensitivity and specificity of test

		Sample_time (days post challenge)							
Host_id	**Challenge**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	**Yes**	2	5	>6	4	3	4	-	-
2	**Yes**	1	-	-	-	4	5	6	5
3	**Yes**	<1	5	4	2	1	-	<1	-
4	**No**	-	-	4	3	-	4	4	-
5	**No**	-	<1	2	-	-	3	-	-
6	**No**	-	-	4	3	-	-	-	<1

- = no growth on selective plate, 1-5 = log concentration ESBL-producing E. coli, <1 is positive but too little to count, >6 positive but too many to count

Infection status of an animal?

- Which infection model:
  - Direct or indirect
  - Latency period
  - Recovery

		Sample_time (days post challenge)							
Host_id	**Challenge**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	**Yes**	2	5	>6	4	3	4	-	-
2	**Yes**	1	-	-	-	4	5	6	5
3	**Yes**	<1	5	4	2	1	-	<1	-
4	**No**	-	-	4	3	-	4	4	-
5	**No**	-	<1	2	-	-	3	-	-
6	**No**	-	-	4	3	-	-	-	<1

- = no growth on selective plate, 1-5 = log concentration ESBL-producing E. coli, <1 is positive but too little to count, >6 positive but too many to count

Infection status of an animal?

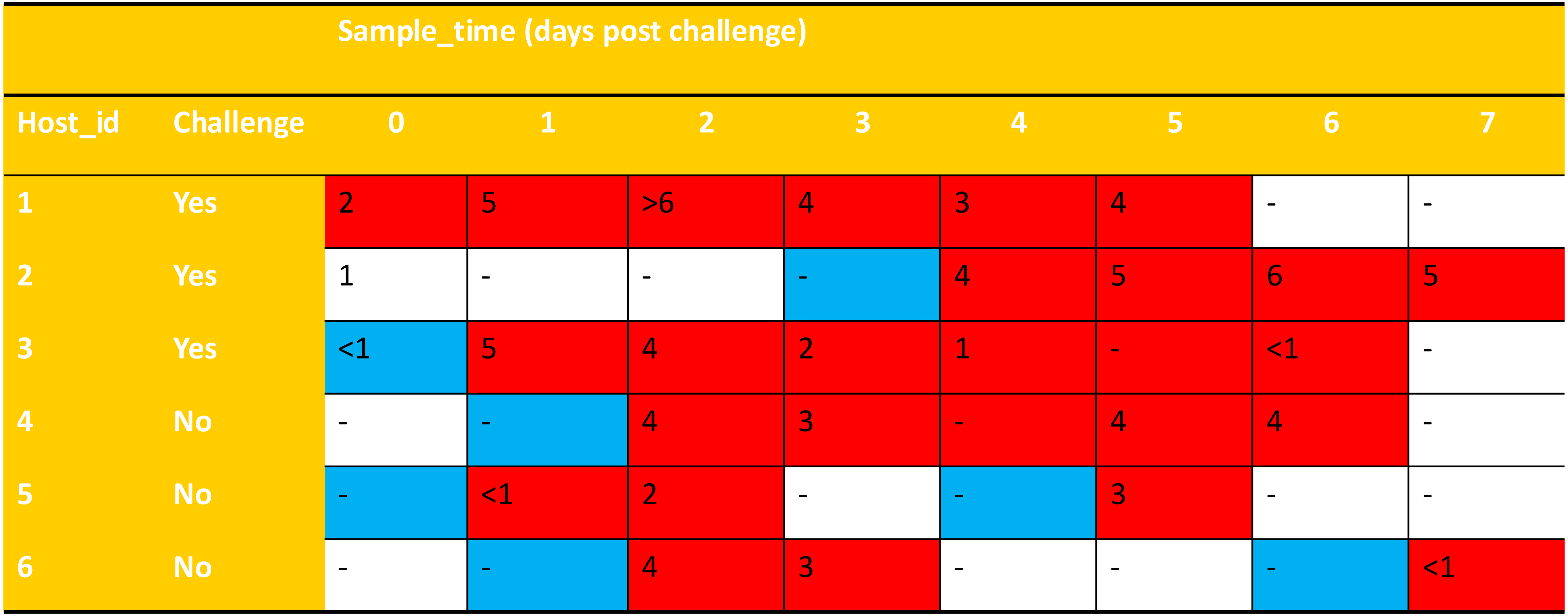
- Example:
  - S-I-S model
  - Low concentration after inoculation = passage

		Sample_time (days post challenge)							
Host_id	**Challenge**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	Yes	2	5	>6	4	3	4	-	-
2	Yes	1	-	-	-	4	5	6	5
3	Yes	<1	5	4	2	1	-	<1	-
4	No	-	-	4	3	-	4	4	-
5	No	-	<1	2	-	-	3	-	-
6	No	-	-	4	3	-	-	-	<1

- = no growth on selective plate, 1-5 = log concentration ESBL-producing E. coli, <1 is positive but too little to count, >6 positive but too many to count

Infection status of an animal?

- Example:
  - S-L-I-S model
  - Assuming possibility of false negative if only once negative
  - Low concentration after inoculation = passage



		Sample_time (days post challenge)							
Host_id	**Challenge**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	**Yes**	2	5	>6	4	3	4	-	-
2	**Yes**	1	-	-	-	4	5	6	5
3	**Yes**	<1	5	4	2	1	-	<1	-
4	**No**	-	-	4	3	-	4	4	-
5	**No**	-	<1	2	-	-	3	-	-
6	**No**	-	-	4	3	-	-	-	<1

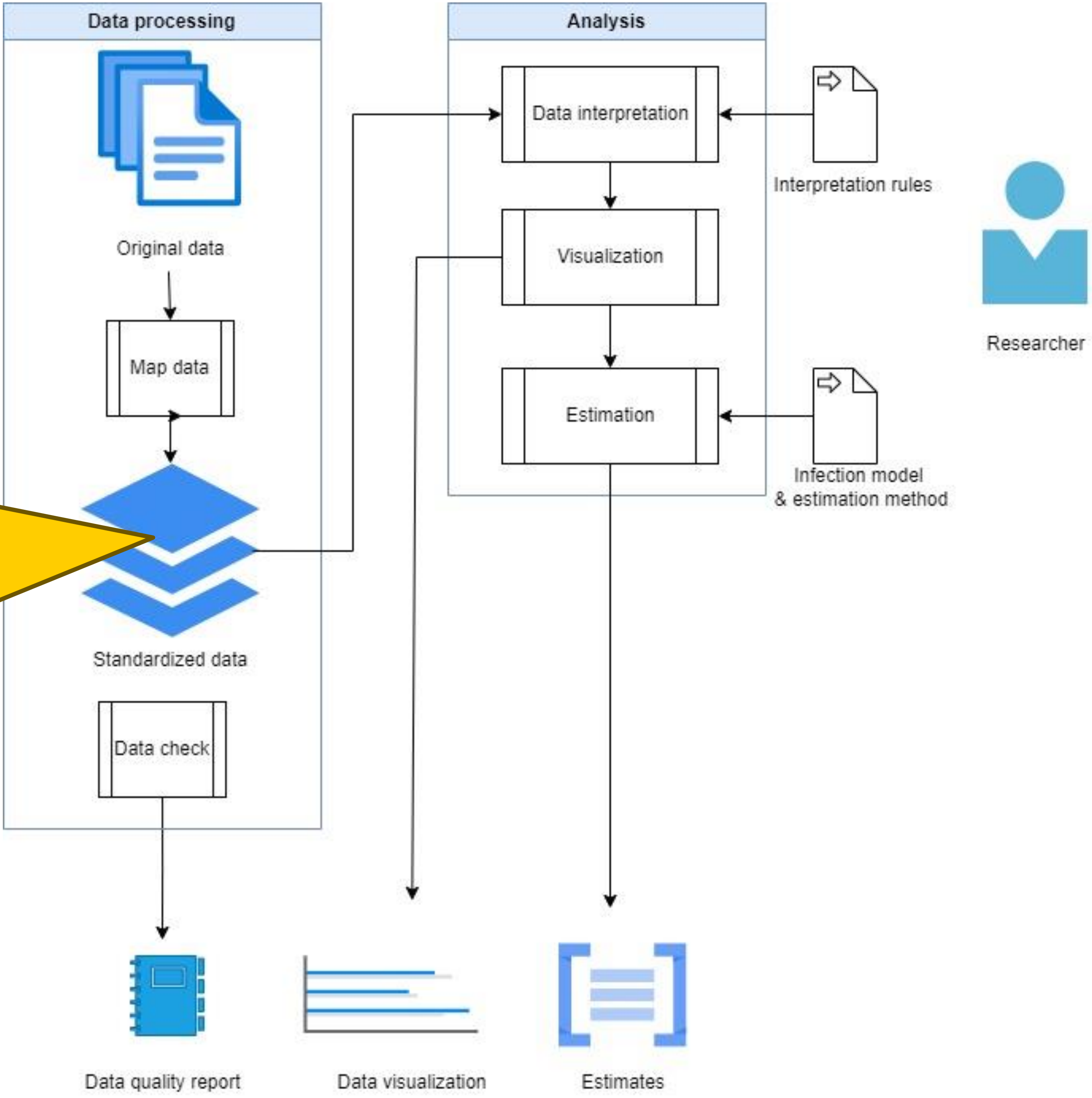
- = no growth on selective plate, 1-5 = log concentration ESBL-producing E. coli, <1 is positive but too little to count, >6 positive but too many to count

Objective of SUMMERFAIR framework

- Facilitate application of different models and assumptions on transmission data in small populations

This requires **standardization of input data**

It allows for
- **Data quality** reports
- **Visualisation**
- Easy documentation and reproducibility of **data interpretation**



Data processing

Original data

Map data

Standardized data

Data check

Analysis

Data interpretation

Visualization

Estimation

Interpretation rules

Infection model
& estimation method

Researcher

Data quality report

Data visualization

Estimates

Based on
Infection Transmission
Ontology

Slavco et al. 2022
https://doi.org/10.1109/eScience55777.2022.00021

Data Mapping

- hosts.csv

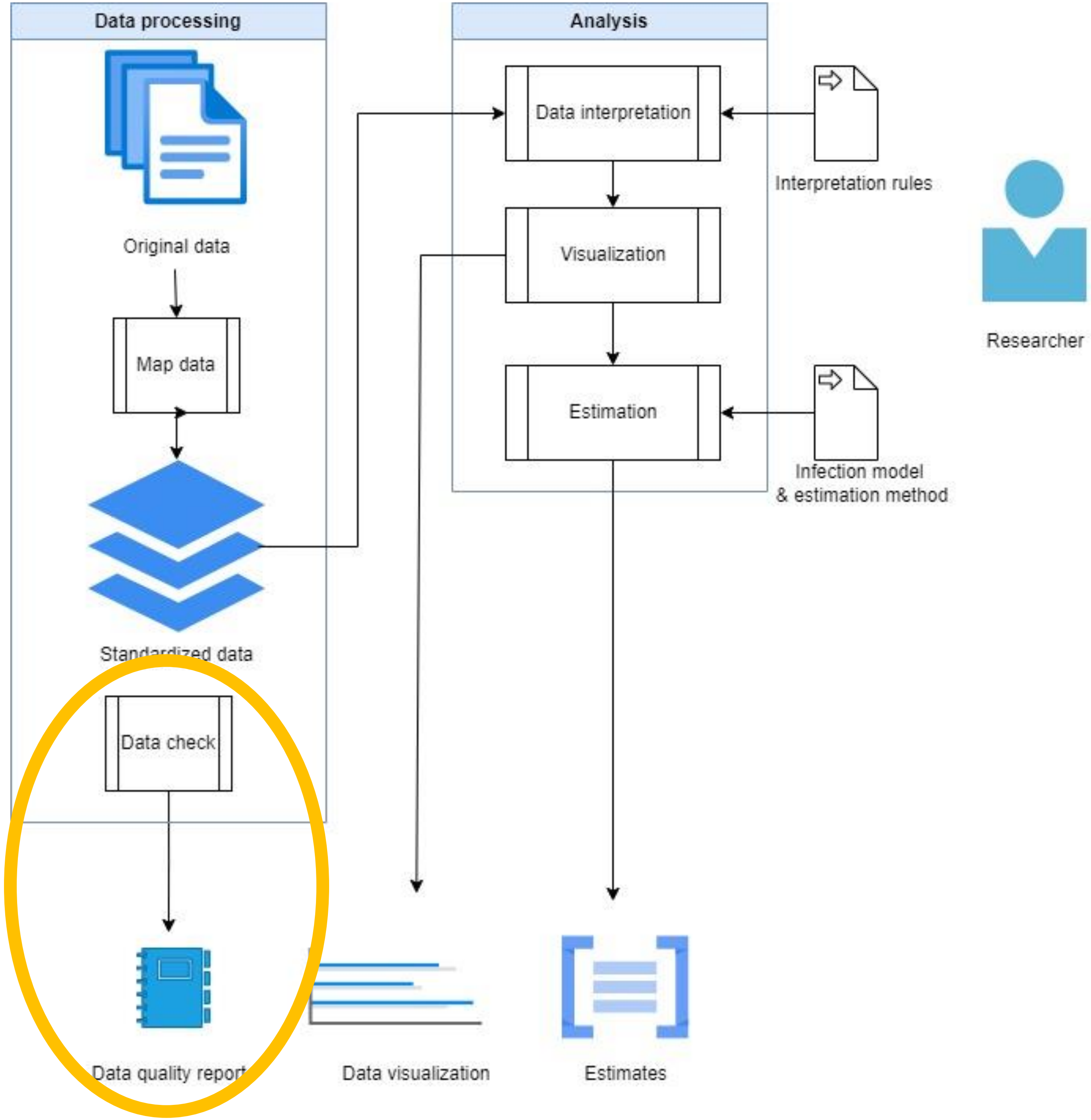
id, sex, group, birth, death

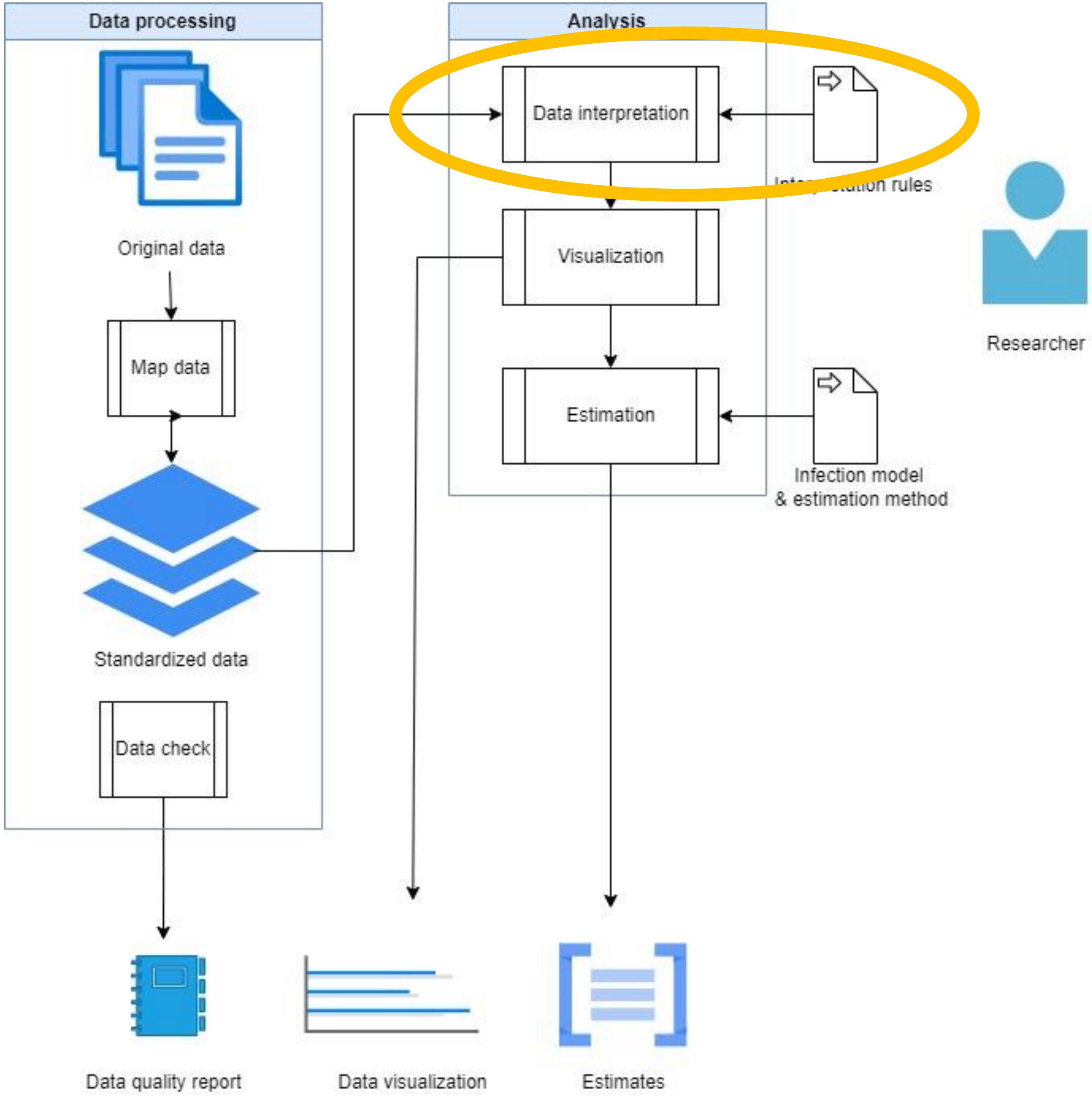
- environments.csv

id, date, time, housing type, housing id, level

- host_events.csv
- environment_events.csv

host id / housing id, date, time, measurement, intervention, treatment


Data processing
Original data
Map data
Standardized data
Data check
Data quality report
Analysis
Data interpretation
Visualization
Estimation
Interpretation rules
Infection model
& estimation method
Researcher
Data visualization
Estimates



Data processing

Original data

Map data

Standardized data

Data check

Analysis

Data interpretation

Interpretation rules

Visualization

Estimation

Infection model
& estimation method

Researcher

Data quality report

Data visualization

Estimates

Infection status of an animal?

- Biological / epidemiological factors:
  - Passage of inoculum
  - Intermittent shedding
  - Sensitivity and specificity of test

		Sample_time (days post challenge)							
Host_id	**Challenge**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	**Yes**	2	5	>6	4	3	4	-	-
2	**Yes**	1	-	-	-	4	5	6	5
3	**Yes**	<1	5	4	2	1	-	<1	-
4	**No**	-	-	4	3	-	4	4	-
5	**No**	-	<1	2	-	-	3	-	-
6	**No**	-	-	4	3	-	-	-	<1

- = no growth on selective plate, 1-5 = log concentration ESBL-producing E. coli, <1 is positive but too little to count, >6 positive but too many to count

Data interpretation rules

- Generic structure:

```
##generic of a rule ####
rule.generic <-function(timeseries,var.id,…){…}
```

- Returns a vector with 0,1,2 or 3 which codes for susceptible, latent, infectious and recovered.

- Preprogrammed *rules* are names
  rule.<interpretation>.<additional features>

Data interpretation rules

- Preprogrammed rules with one input vector

- `asis, sincefirst, consecutive`

- Preprogrammed rules with multiple input vectors

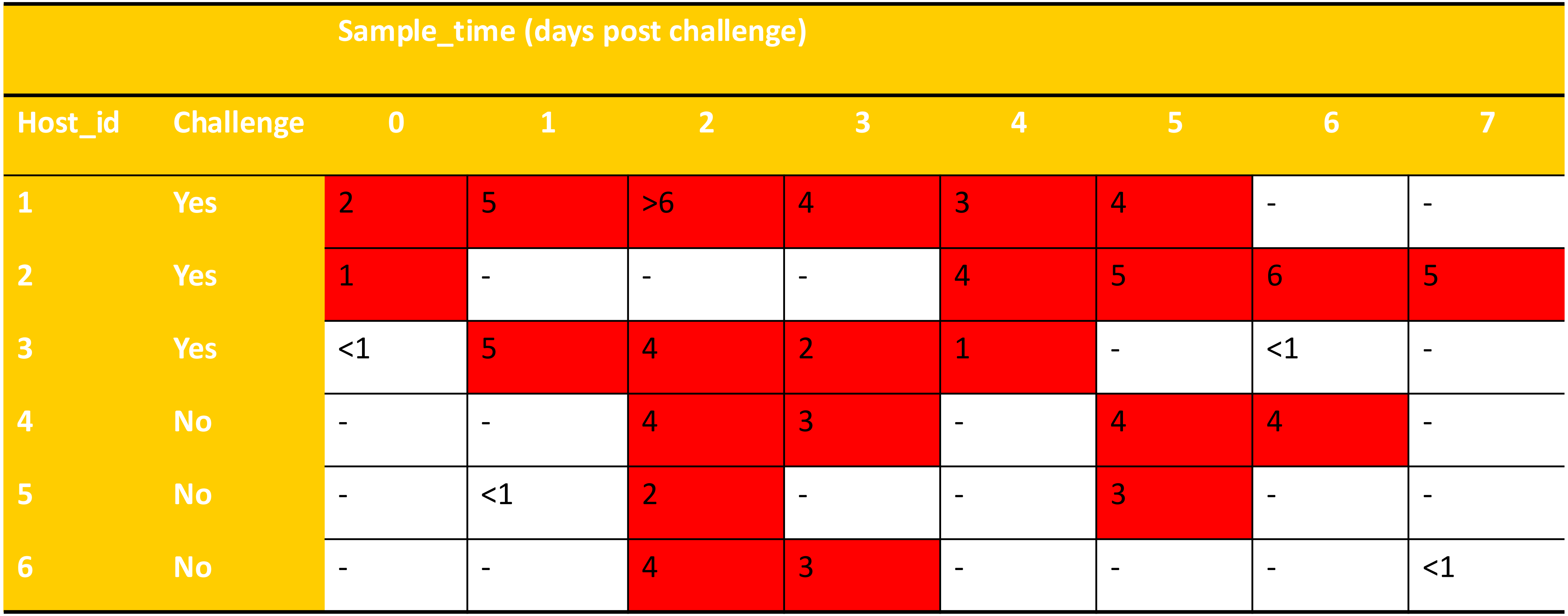
- `any, sinceany, infectiousrecovered`

- Additional features:

- `recode, detectionLimit, cutoff`

Interpretation rule `asis.cutoff`

`cutoff = 1`



		Sample_time (days post challenge)							
Host_id	**Challenge**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	Yes	2	5	>6	4	3	4	-	-
2	Yes	1	-	-	-	4	5	6	5
3	Yes	<1	5	4	2	1	-	<1	-
4	No	-	-	4	3	-	4	4	-
5	No	-	<1	2	-	-	3	-	-
6	No	-	-	4	3	-	-	-	<1

- = no growth on selective plate, 1-5 = log concentration ESBL-producing E. coli, <1 is positive but too little to count, >6 positive but too many to count

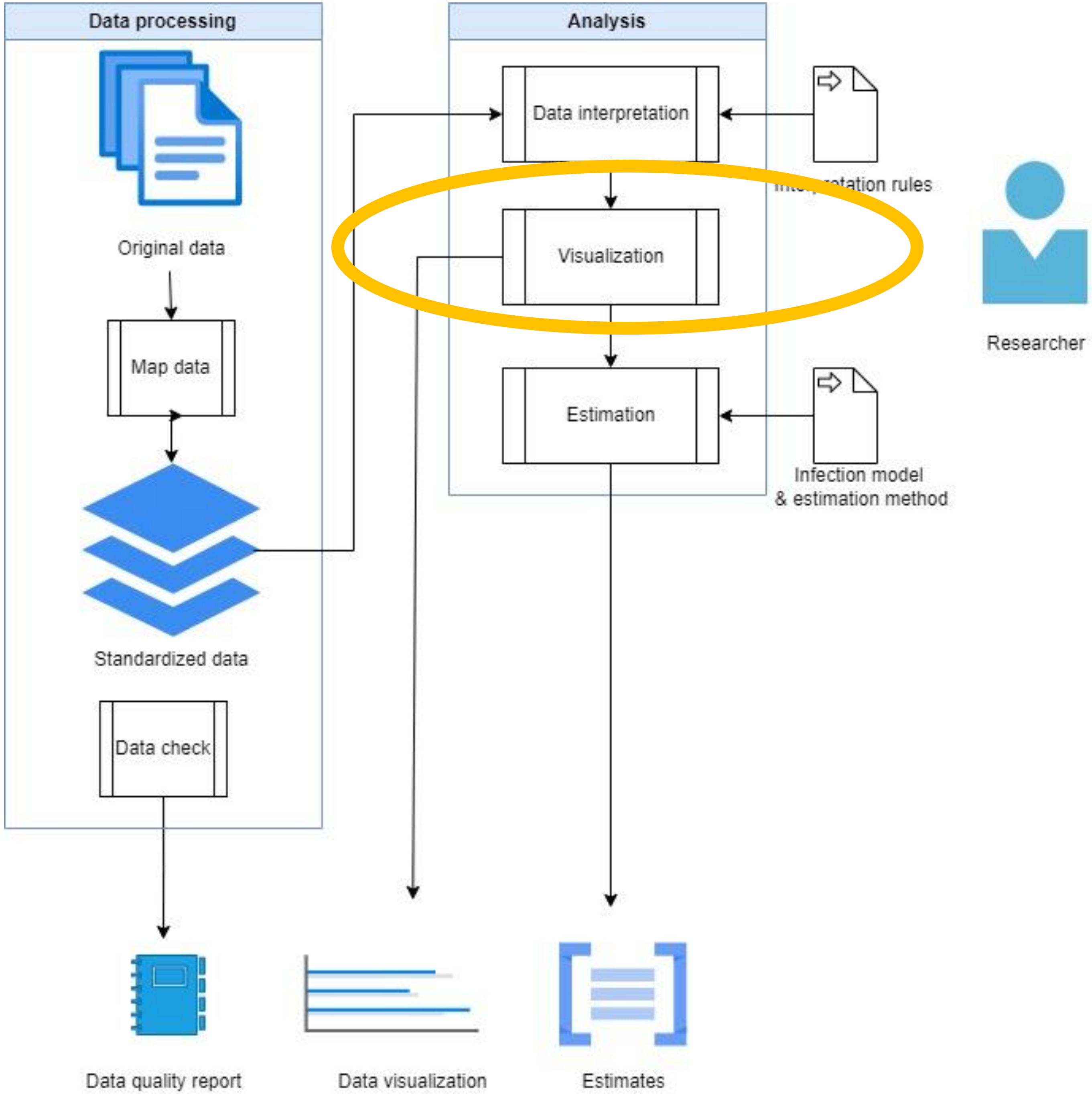
Interpretation rule `consecutive.cutoff`

`cutoff = 1`

`2 consecutive days  = positive / negative`

		Sample_time (days post challenge)							
Host_id	**Challenge**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	Yes	2	5	>6	4	3	4	-	-
2	Yes	1	-	-	-	4	5	6	5
3	Yes	<1	5	4	2	1	-	<1	-
4	No	-	-	4	3	-	4	4	-
5	No	-	<1	2	-	-	3	-	-
6	No	-	-	4	3	-	-	-	<1

- = no growth on selective plate, 1-5 = log concentration ESBL-producing E. coli, <1 is positive but too little to count, >6 positive but too many to count



Data processing

Original data

Map data

Standardized data

Data check

Analysis

Data interpretation

Visualization

Estimation

Interpretation rules

Infection model
& estimation method

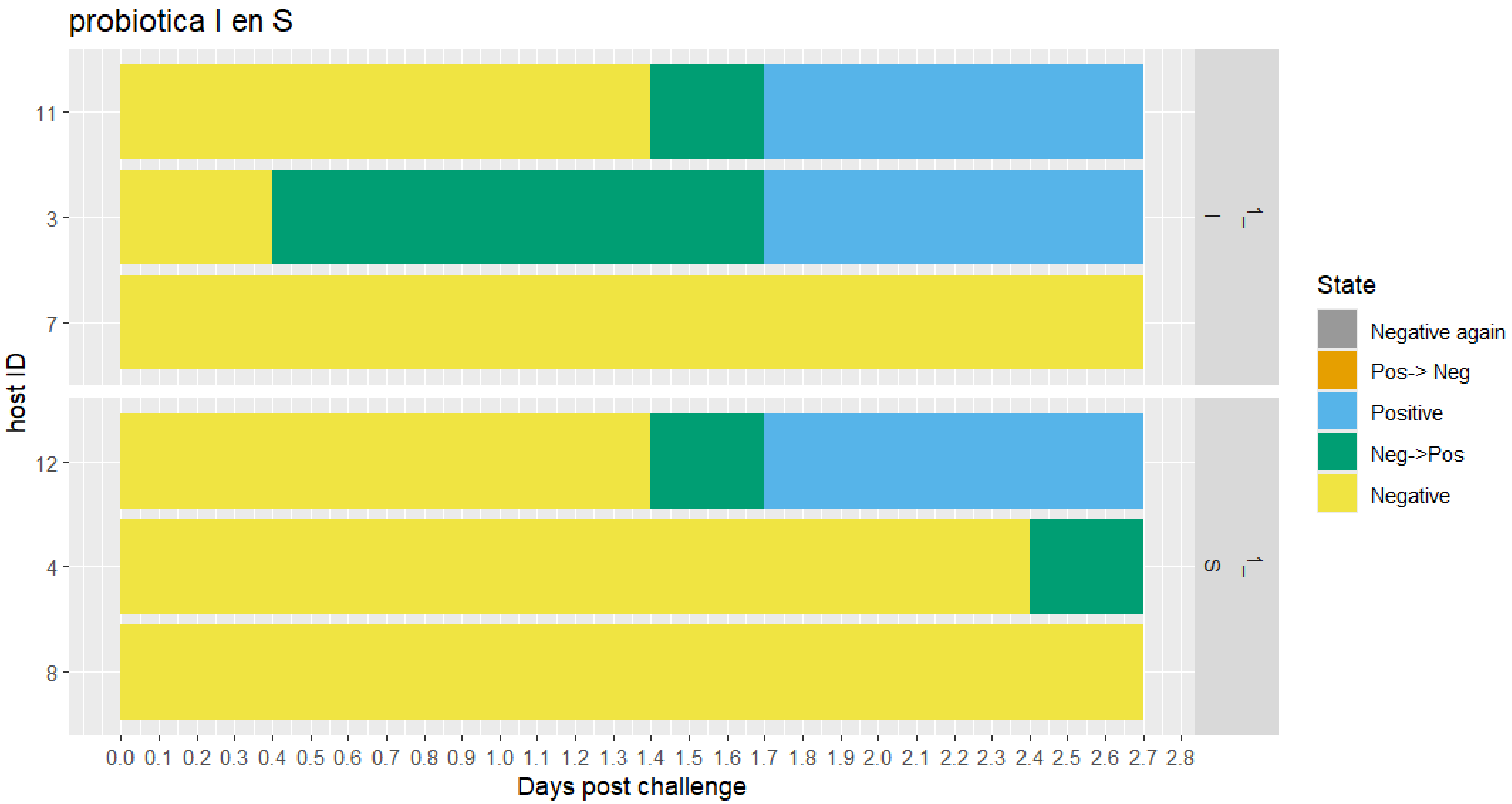
Researcher

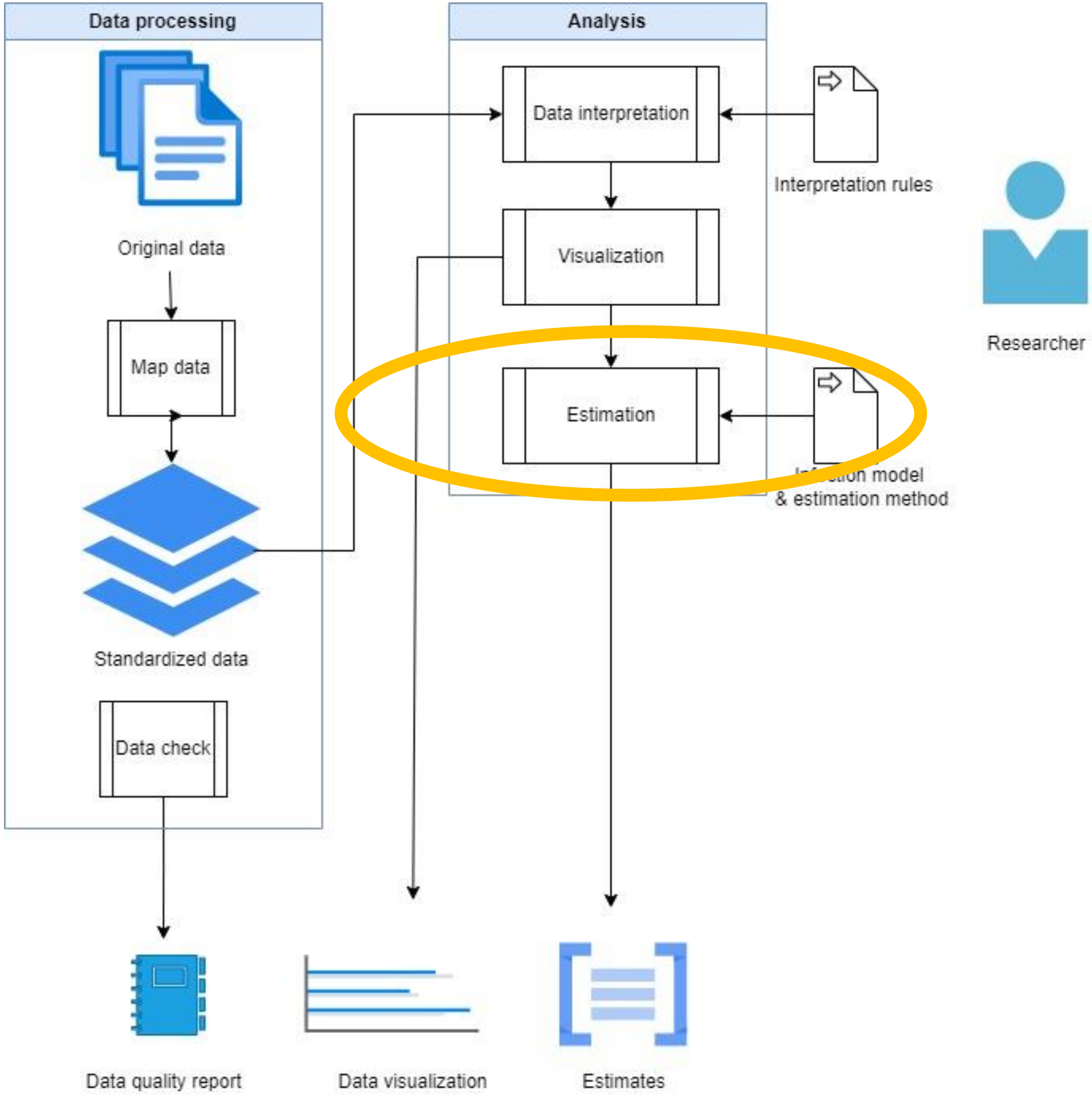
Data quality report

Data visualization

Estimates

Data visualisation





Data processing

Original data

Map data

Standardized data

Data check

Analysis

Data interpretation

Visualization

Estimation

Interpretation rules

Infection model
& estimation method

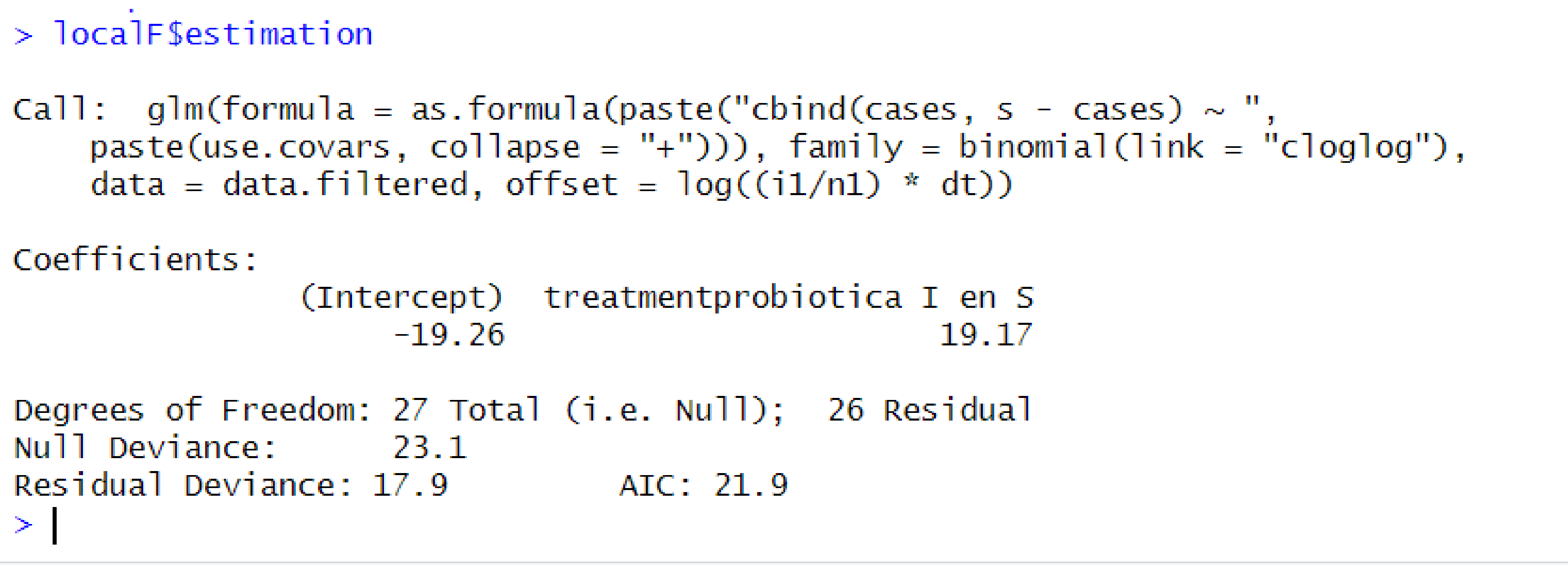
Researcher

Data quality report

Data visualization

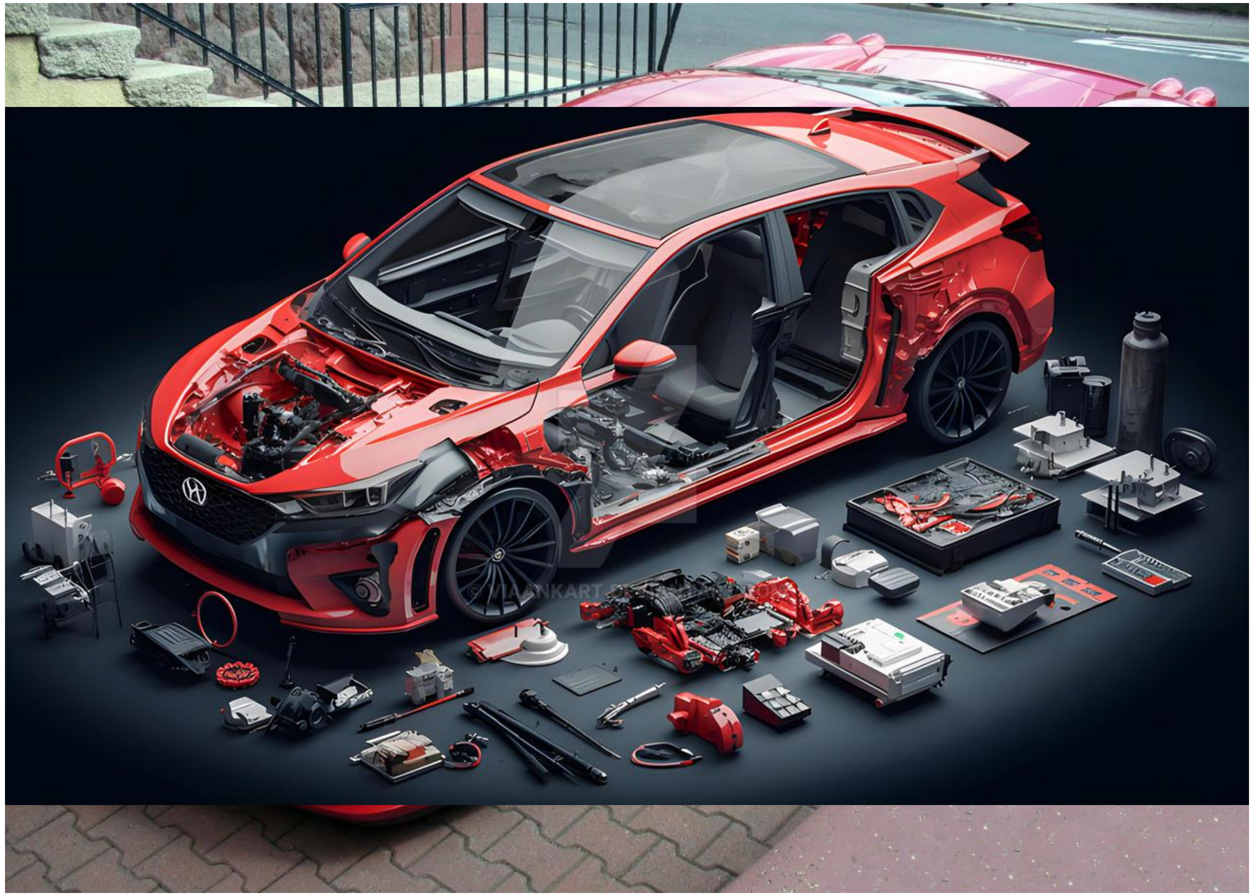
Estimates

Estimation



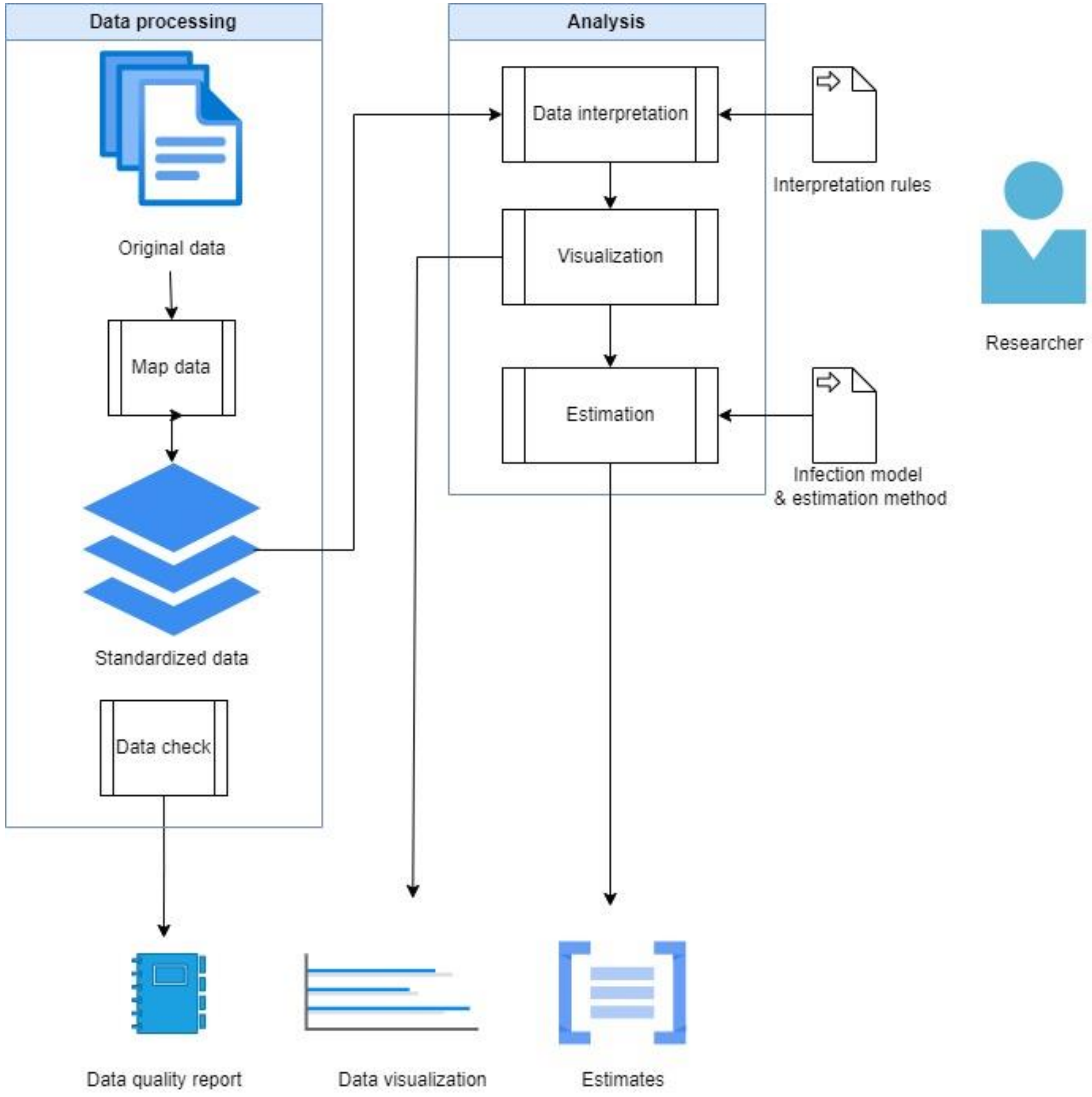
```
> localF$estimation

Call:  glm(formula = as.formula(paste("cbind(cases, s - cases) ~ ",
    paste(use.covars, collapse = "+"))), family = binomial(link = "cloglog"),
    data = data.filtered, offset = log((i1/n1) * dt))

Coefficients:
               (Intercept)  treatmentprobiotica I en S
                    -19.26                       19.17

Degrees of Freedom: 27 Total (i.e. Null);  26 Residual
Null Deviance:      23.1
Residual Deviance: 17.9        AIC: 21.9
> |
```

Work in progress



Discussion

- Data standardization
- Data interpretation
    - Automated
    - Documented
    - Reproducible
- Estimation
    - Reproducible
    - Reusable
- Adaptable code



Universiteit
Utrecht

Sharing science,
shaping tomorrow

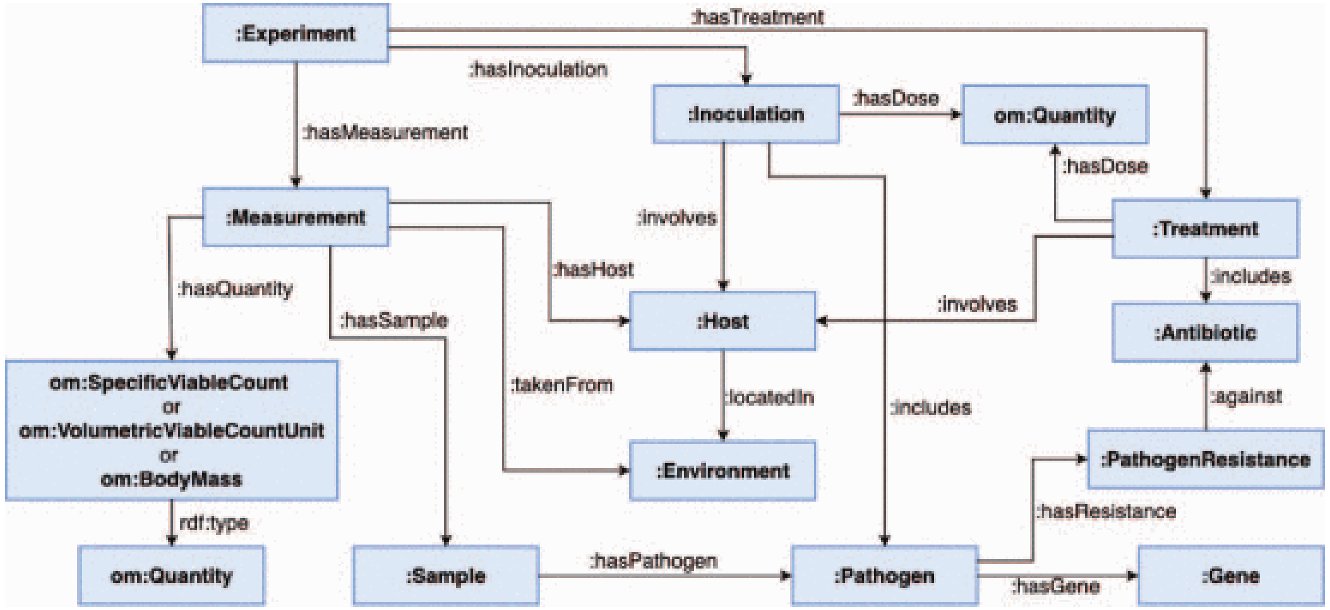
We thank Christine Staiger, Raoul Schram and Stefano Rapisarda, the research engineering group.

We also acknowlegde the work of Elena Slavco (https://doi.org/10.1109/eScience55777.2022.00021)

This work was funded by ZonMW (grant nr. 50-54100-98-408) and FAIR innovation fund of the Utrecht University

https://github.com/UtrechtUniversity/autumn-fair/blob/main/docs/SummerFAIR_schema_proposal_V04.pdf

Data Mapping



Infection Transmission Ontology
Slavco et al. 2022
https://doi.org/10.1109/eScience55777.2022.00021