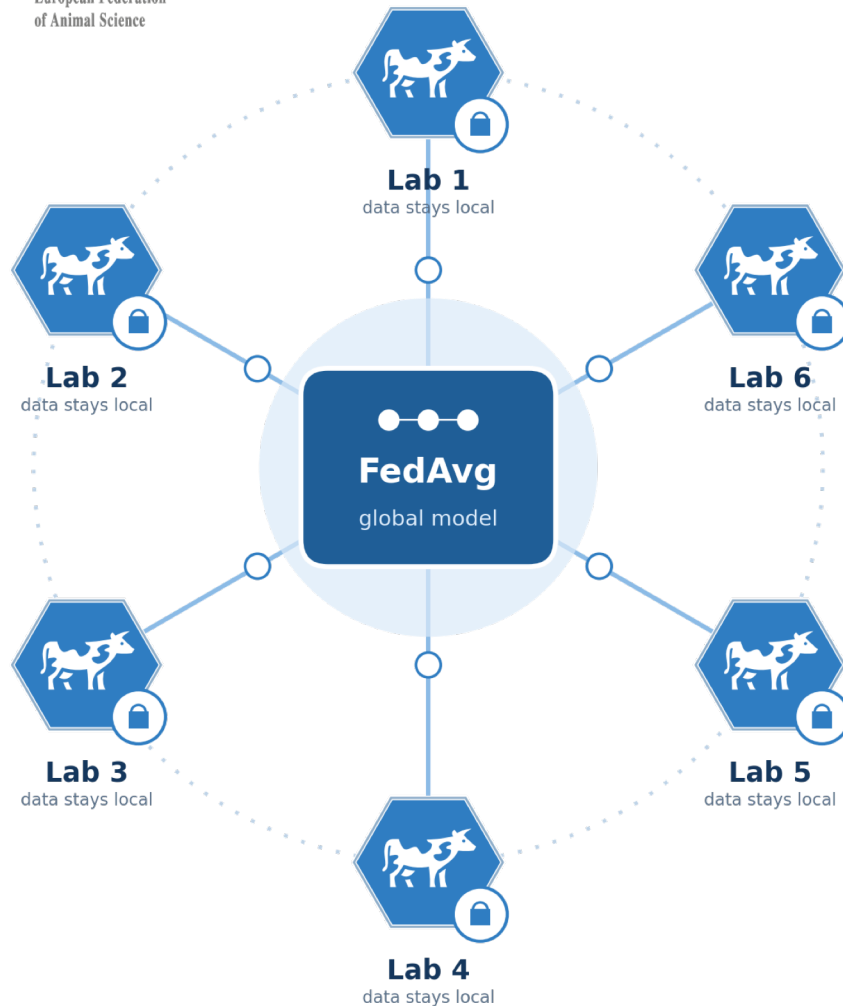




Federated Learning for Bovine Respiratory Disease prediction Across European Veterinary laboratories

The data never leaves. The intelligence does.

S. Noor · J. Degroote · B. Pardon · G. Van Schaik · M. Hostens

Ghent University · Utrecht University · Cornell University

The Cattle Barometer today

Rich history - but it only looks backward

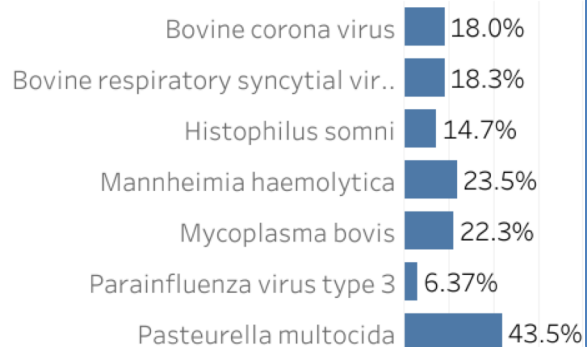
Scan to explore
the live tool



Number/Percentage positive samples

22.3%

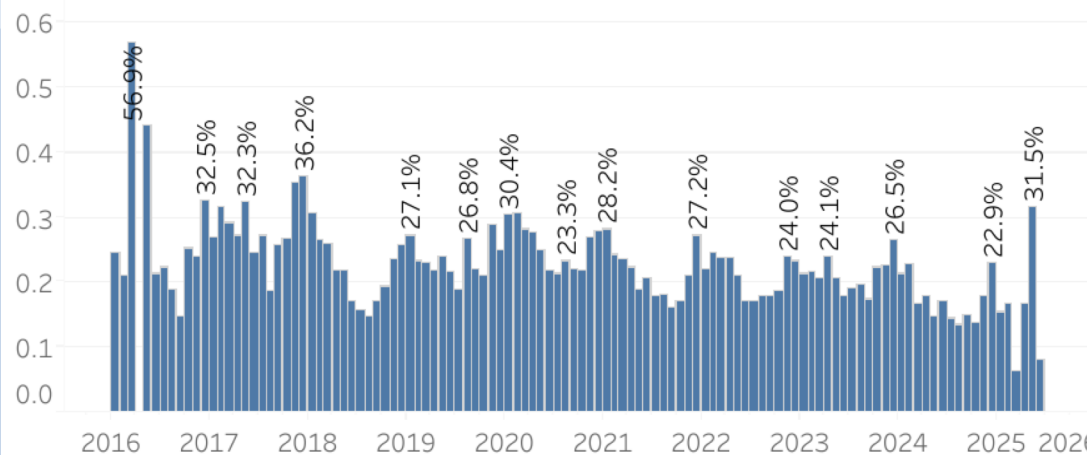
By pathogen



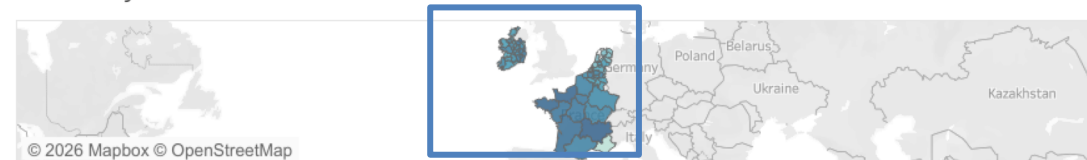
Date

All values

Number/Percentage positive samples over time



Country



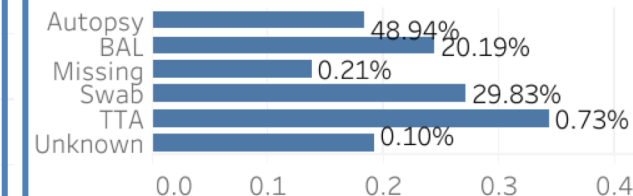
Number of positives or positive sample rate?

Positive sample rate (%)

Country

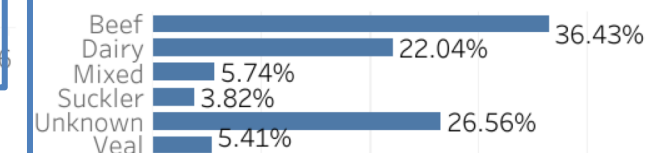
All

By sample type

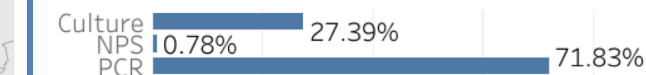


Value

By production type

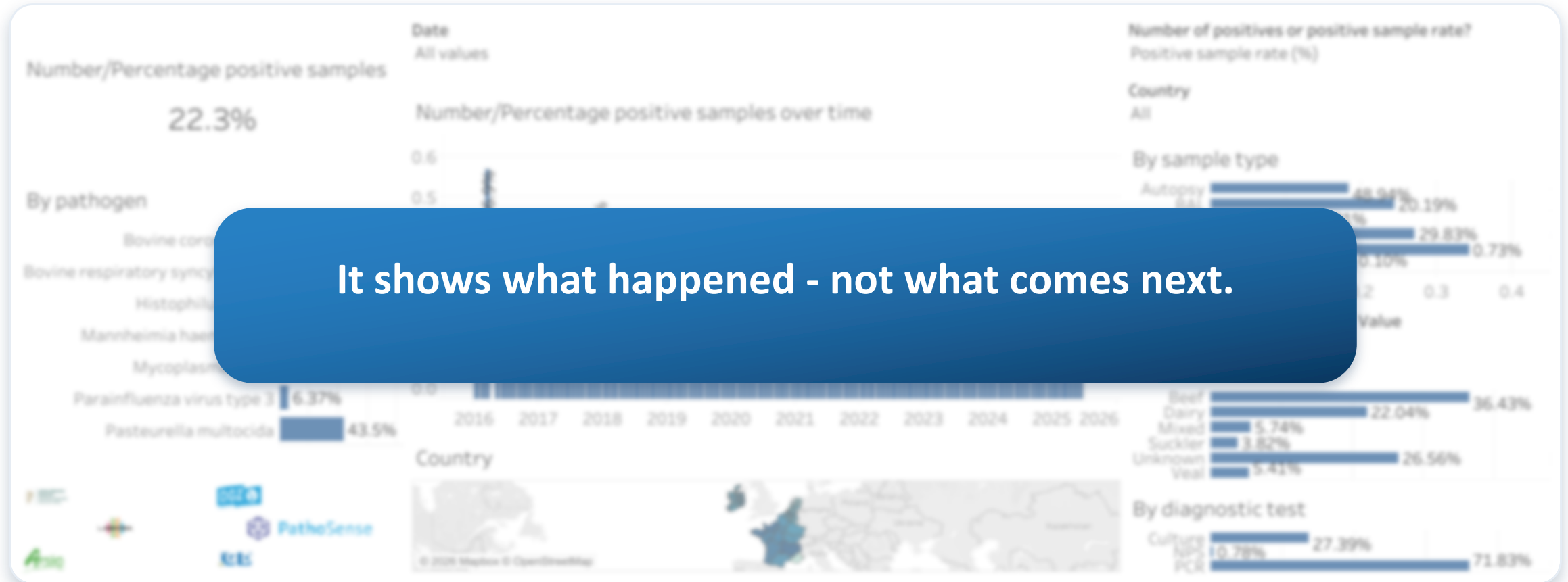


By diagnostic test



The Cattle Barometer today

Rich history - but it only looks backward



Scan to explore
the live tool



The problem — and why it matters

Bovine Respiratory Disease is the biggest health burden in dairy and beef:



Welfare

Sick, suffering calves and feedlot cattle



Mortality & cost

Deaths, costly treatment, lost growth



Antibiotics & AMR

BRD drives heavy antibiotic use

But predicting BRD needs data from many labs — and privacy law blocks pooling. Labs even record different pathogens and tests, so their feature spaces don't match.



Can we predict BRD across Europe — without any lab ever sharing its raw data?

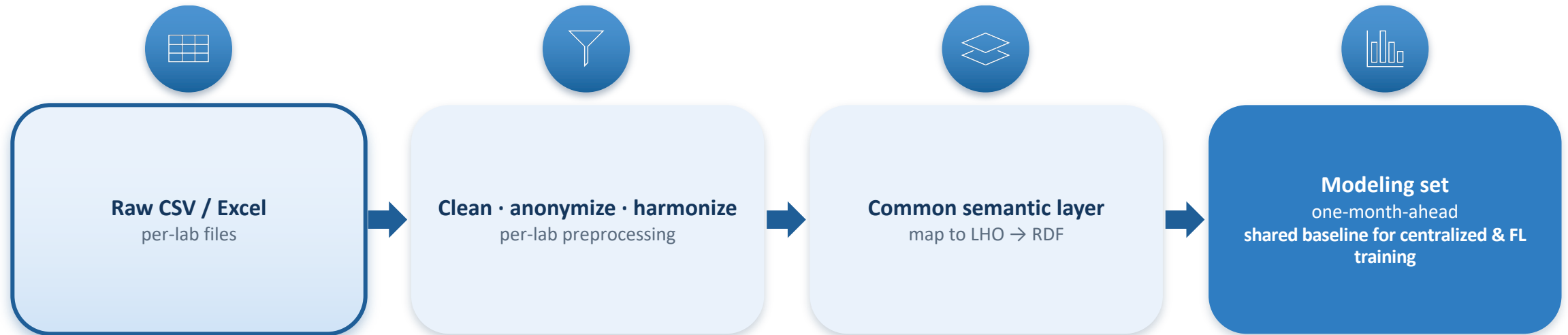
The data — Cattle Barometer

 **164,814**

diagnostic records · 2016–2025

6 laboratories · 4 countries (BE · NL · IE · FR) · 7 BRD pathogens · 1-month-ahead, monthly

DATA PIPELINE: the same harmonized baseline feeds the model — for both centralized & FL training



Methodology: federated learning on Solid PODs

Instead of pooling raw data in one place, we send the model to each lab's Solid POD — raw records never leave.

Centralized training



Data



Model

All data is collected into one place. Sensitive records must leave the owner.



blocked by privacy law

Federated learning · Solid PODs



Model



PODs (Data)

The model travels to each lab's Solid POD, trains locally — only the learned weights return. Raw data never leaves.



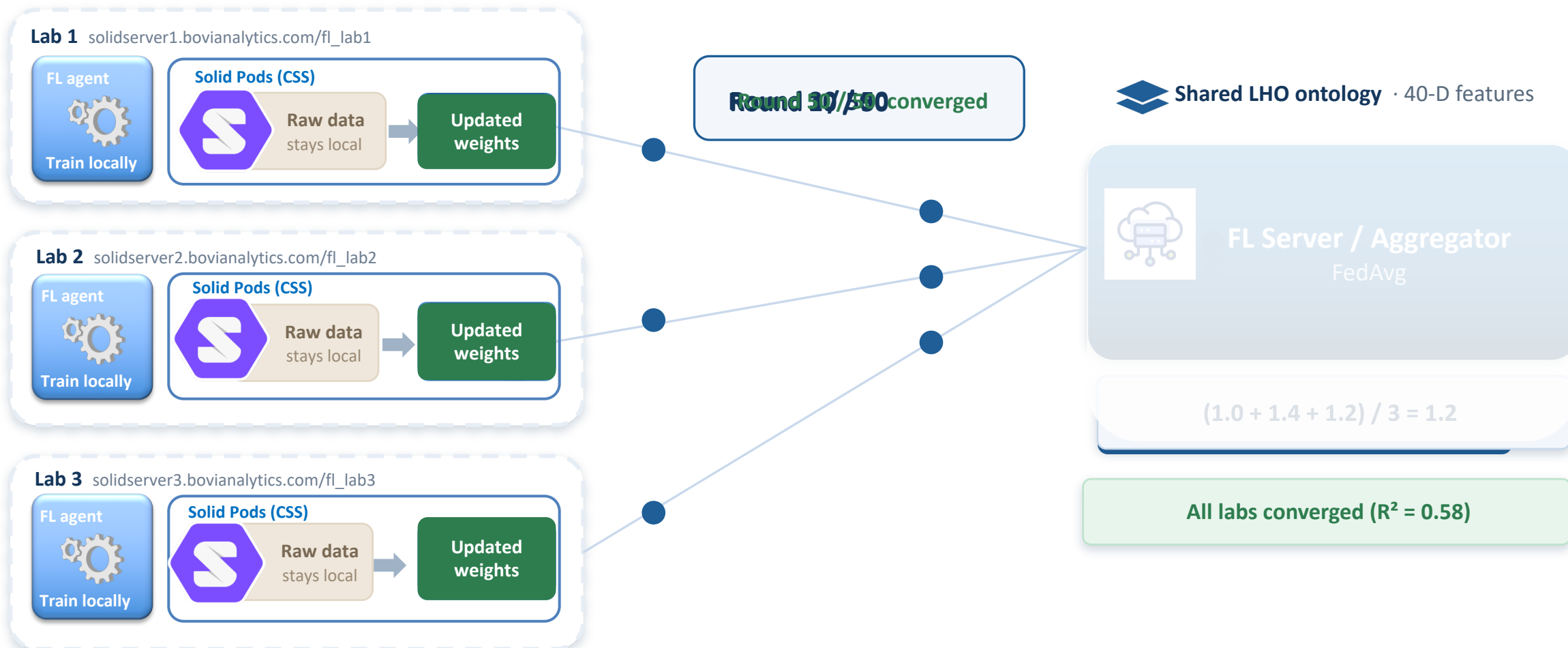
privacy preserved



Solid (Social Linked Data) — a decentralized-web standard from Sir Tim Berners-Lee (MIT, since 2016). Each lab keeps its records in its own POD and the model comes to the POD. **Applying FL-on-Solid to livestock diagnostics is our contribution to animal science.**

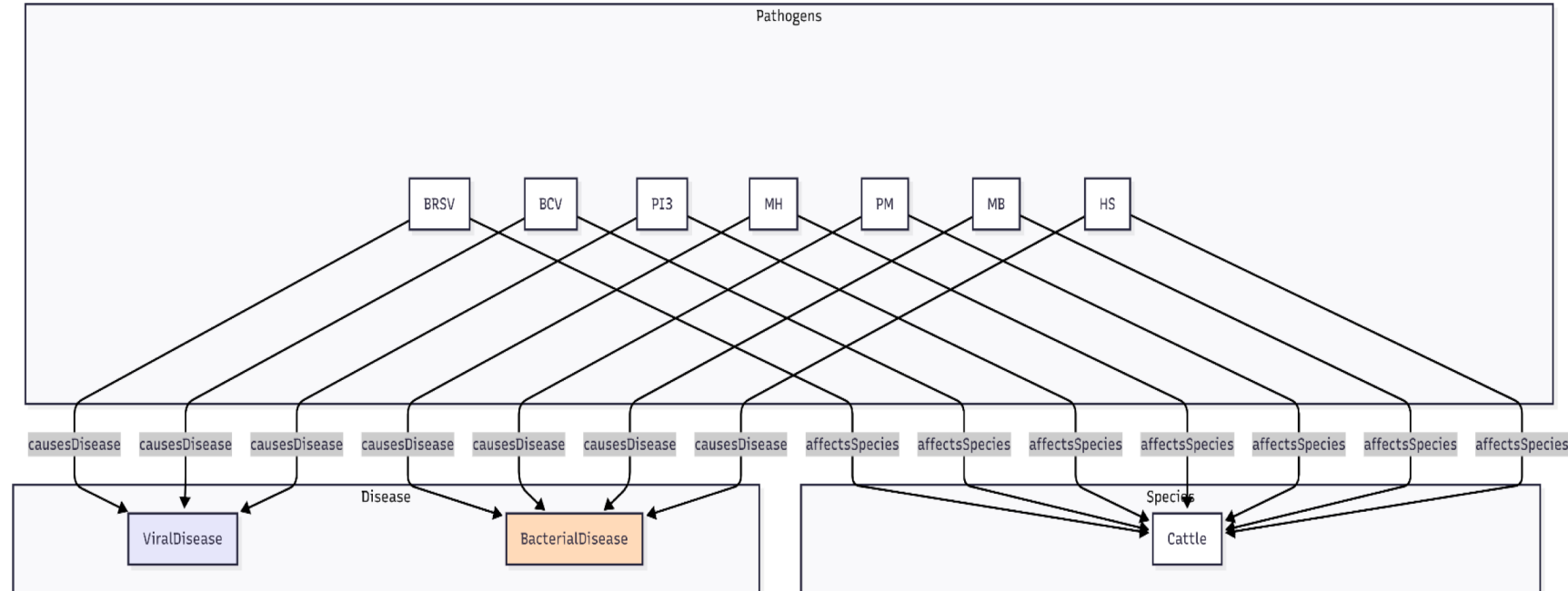
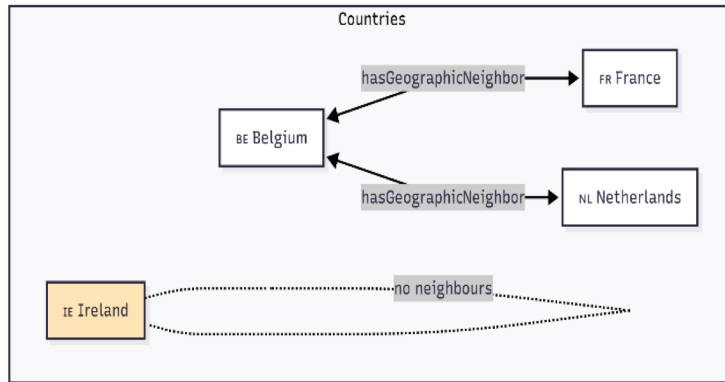
Federated learning in action

FedAvg over Solid PODs — weights travel to the server; raw data stays locked in each lab.



1 · Train locally 2 · Sending weights to the server 3 · Aggregating weights 4 · Sending weights back to the server 5 · All labs converged ($R^2 = 0.58$) 6 · Data never leaves

The foundation: Livestock Health Ontology (LHO)



Shared semantic model

One structural vocabulary across all labs



Interoperable & FAIR

Lab results become machine-readable



Linked to standard vocabularies

Reuses AGROVOC (FAO), AHOL, ANDO, NCBI Taxonomy & NCIT



Published & openly available

On [BioPortal](#) & [AgroPortal](#) ; code on GitHub, archived on Zenodo

Scan to explore it live



Browse live - AgroPortal



Source code - GitHub

Without a shared vocabulary, labs with different pathogens cannot be compared or combined.

Ontology solves the feature-mismatch problem



Without ontology

Lab 1: BCV · BRSV · PI3 · MH

Lab 2: BCV · HS · MB

Lab 3: BRSV · PM · MH

Different dimensions → models can't be averaged ✗



With LHO ontology

All labs query the ontology: BCV · BRSV · PI3 · MH · PM · MB · HS

Every lab uses the same 7 pathogens: Bovine coronavirus (BCV), Bovine respiratory syncytial virus (BRSV), parainfluenza virus type 3 (PI3), Mannheimia haemolytica (MH), Pasteurella (PM), Mycoplasma bovis (MB), and Histophilus somni (HS)

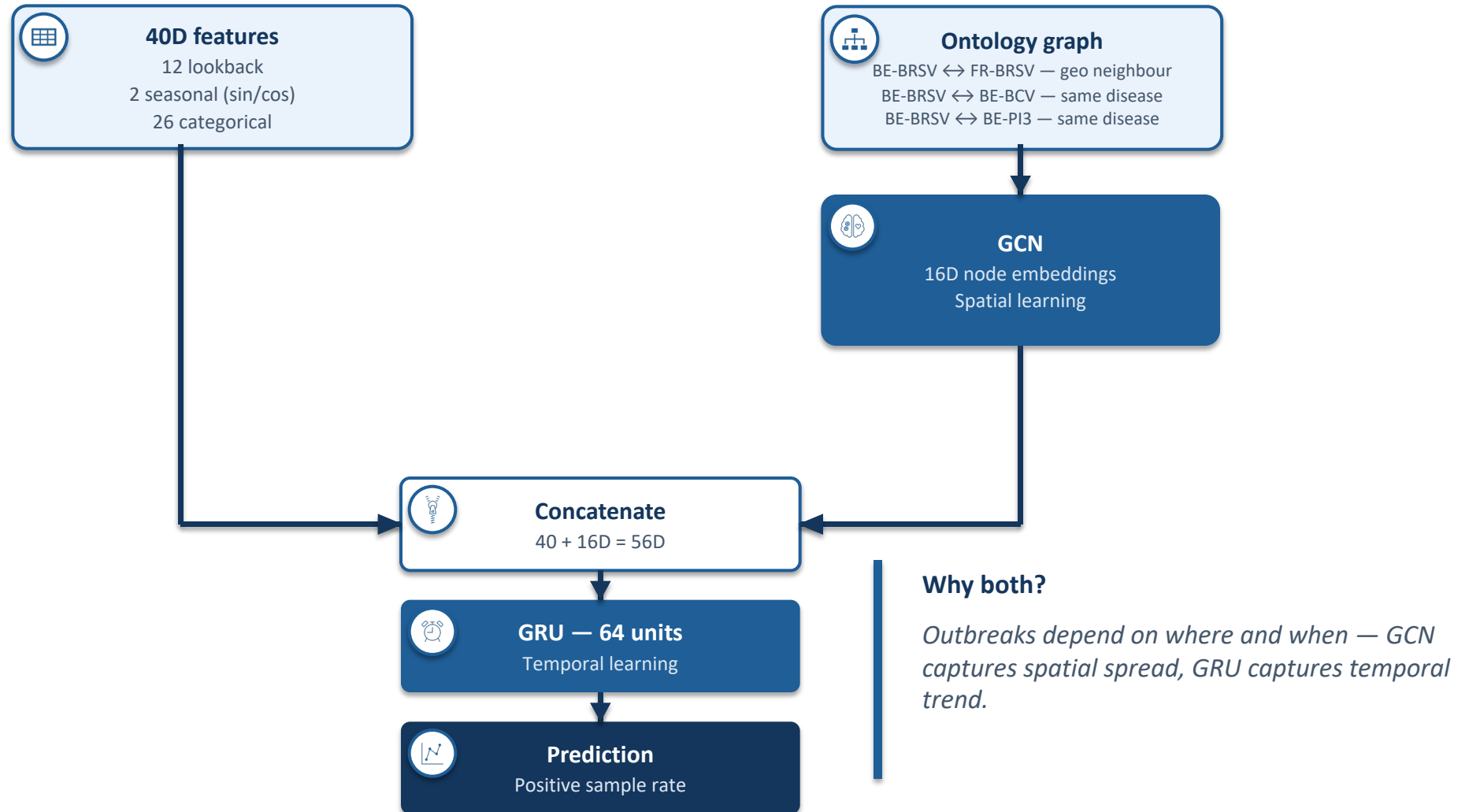
(zero for any it doesn't have)

Same dimensions → FedAvg works ✓

Queried via SPARQL — extract categories & relationships

The model: GCN + GRU

Graph Convolutional Network (GCN) + Gated Recurrent Unit (GRU)



Experimental setup

From data to federated learning — two model variants evaluated on one shared protocol.



MODULE 1

GCN-GRU Centralized

RDF data + LHO ontology + Semantic knowledge

All labs pooled · best-case benchmark



MODULE 2

GCN-GRU Federated

RDF data + LHO ontology + Privacy on Solid PODs

Trains locally on each Solid POD · FedAvg



Common evaluation protocol across both modules



60 / 20 / 20
data split

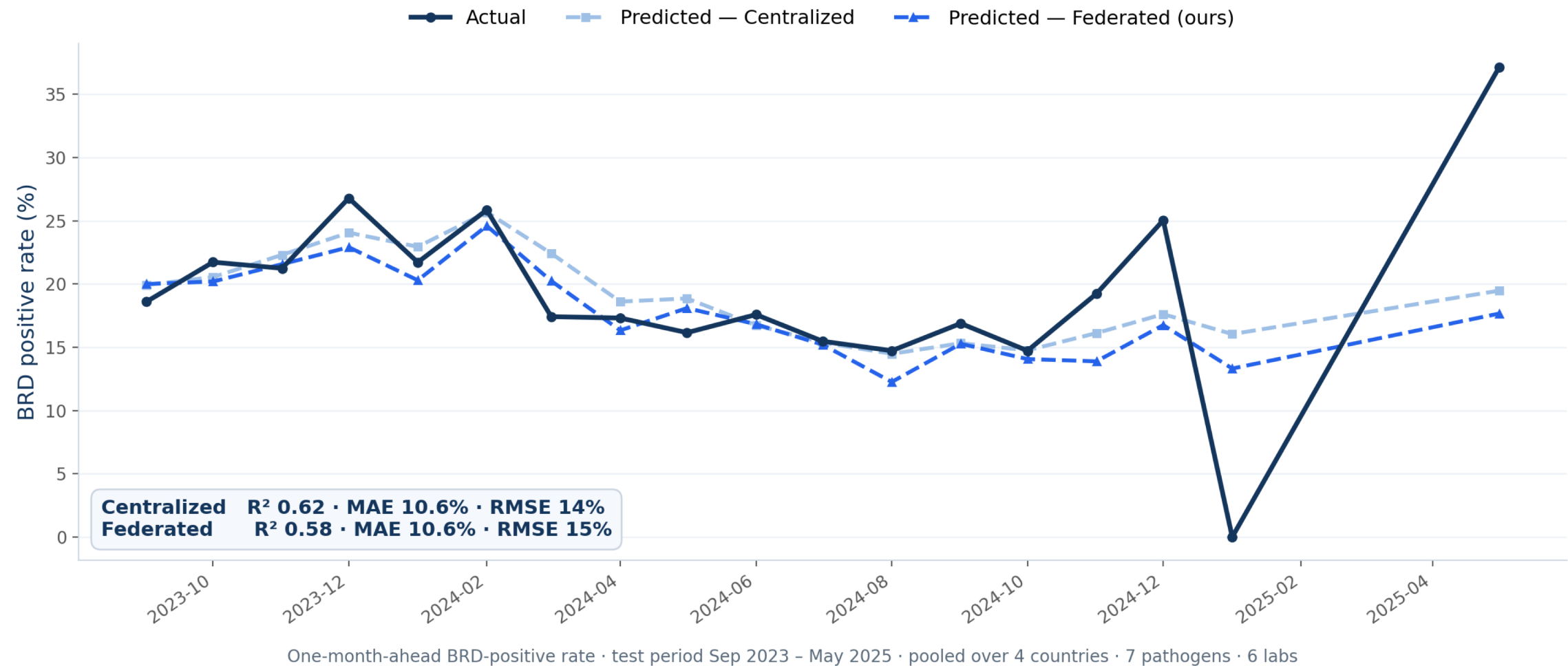


Metrics:
 R^2 , RMSE, MAE



BRD positive sample rate,
one month ahead prediction

Results — actual vs predicted: Centralized vs Federated



Centralized vs Federated — minimal loss

Metric	Centralized	Federated	Trade-off
R^2	0.62	0.58	slight
MAE	10.6%	10.6%	identical
RMSE	14%	15%	marginal

Identical MAE confirms equivalent average accuracy. The model keeps all data local — privacy at almost no cost in accuracy.

Live system — federated learning on Solid PODs

The image displays a live federated learning system interface. On the left, a Windows PowerShell terminal window shows logs from various lab servers (sidecar-lab4-1, fl-server-1, sidecar-lab5-1, sidecar-lab6-1, sidecar-lab2-1, sidecar-lab5-1) and a central fl-server-1. The logs indicate the progression of federated learning rounds, including weight pulling, aggregation, and global model validation. Key metrics like R2 values and sample counts are visible. On the right, a web browser displays the 'DECIDE FL Federation Control Panel' at localhost:8090. The dashboard provides a summary of the federated model performance, including ontology-driven horizontal federated learning (GCN + FedAvg) results. It shows the best global val R2 (target ~0.5844), current val R2, val MAE, and val RMSE. The number of rounds completed (5) and weight submissions (5) are also displayed. Below the performance metrics, the dashboard lists participating labs (Lab 1 and Lab 2) with their FL participation status (ACTIVE) and options to leave the federation or view pod data and FL config. A video player in the bottom right corner shows a recording of the system's operation.

Windows PowerShell

```
sidecar-lab4-1 | 2026-06-09 00:47:40 [sidecar] INFO [Lab 4] Weights published to pod
fl-server-1 | 2026-06-09 00:47:40,731 INFO [AUTH] Access token obtained from http
fl-server-1 | 2026-06-09 00:47:41,212 INFO Lab 5: pulled weights from pod (http://
les=639 | fed val R2=-2.4262 (best -2.4262)
sidecar-lab5-1 | 2026-06-09 00:47:41 [sidecar] INFO [Lab 5] Server pulled weights fr
sidecar-lab5-1 | 2026-06-09 00:47:41 [sidecar] INFO [Lab 5] Round 1 finished (this l
sidecar-lab5-1 | 2026-06-09 00:47:41 [sidecar] INFO
sidecar-lab5-1 | 2026-06-09 00:47:41 [sidecar] INFO [Lab 5] Round 1: global model va
fl-server-1 | 2026-06-09 00:47:41,472 INFO Lab 4: pulled weights from pod (http://
es=653 | fed val R2=-2.4139 (best -2.4139)
sidecar-lab4-1 | 2026-06-09 00:47:41 [sidecar] INFO [Lab 4] Server pulled weights fr
sidecar-lab4-1 | 2026-06-09 00:47:41 [sidecar] INFO [Lab 4] Round 6 finished (this l
sidecar-lab4-1 | 2026-06-09 00:47:41 [sidecar] INFO
sidecar-lab4-1 | 2026-06-09 00:47:41 [sidecar] INFO [Lab 4] Round 6: global model va
sidecar-lab1-1 | Total samples parsed: 37,535
sidecar-lab1-1 | Date range: 2016-12-02 00:00:00 to 2024-09-26 00:00:00
sidecar-lab1-1 | Positive samples: 9,668 (25.8%)
sidecar-lab1-1 | [Step 2] Aggregating data...
sidecar-lab1-1 | Aggregated records: 8,145
sidecar-lab1-1 | Applying filter: Tested >= 5
sidecar-lab1-1 | Before: 8,145 -> After: 3,005
sidecar-lab1-1 | [Step 3] Checking data availability...
sidecar-lab1-1 | Unique months: 94
sidecar-lab1-1 | Using STANDARD 12-month lookback
sidecar-lab1-1 | [Step 4] Building 40D features...
fl-server-1 | 2026-06-09 00:47:42,057 INFO Lab 6: pulled weights from pod (http://
les=975 | fed val R2=-1.5903 (best -1.5903)
sidecar-lab6-1 | 2026-06-09 00:47:42 [sidecar] INFO [Lab 6] Server pulled weights fr
sidecar-lab6-1 | 2026-06-09 00:47:42 [sidecar] INFO [Lab 6] Round 1 finished (this l
sidecar-lab6-1 | 2026-06-09 00:47:42 [sidecar] INFO
sidecar-lab6-1 | 2026-06-09 00:47:42 [sidecar] INFO [Lab 6] Round 1: global model va
sidecar-lab2-1 | 2026-06-09 00:47:42 [sidecar] INFO [Lab 2] === Async round 2/50 ===
sidecar-lab2-1 | 2026-06-09 00:47:42 [sidecar] INFO [Lab 2] Global model val R2 (bef
sidecar-lab2-1 | 2026-06-09 00:47:43 [sidecar] INFO [Lab 2] Training complete - Val
sidecar-lab5-1 | 2026-06-09 00:47:43 [sidecar] INFO [Lab 5] === Async round 2/50 ===
```

DECIDE FL Federation Control Panel

A federated learning system where six laboratories across four countries train a shared disease-prediction model without ever sharing raw data, demonstrated using the Cattle Barometer as a testbed.

Federated Model Performance
Ontology-Driven Horizontal Federated Learning (GCN + FedAvg)

Metric	Value
best global val R ² (target ~0.5844)	-7.7420
current val R ²	-7.7420
val MAE	0.3482
val RMSE	0.3695

5 rounds completed (max across labs) | 5 weight submissions

14 training samples (per round) | 6/6 labs participating

Federated evaluation - each lab scores the shared global model on its own held-out data; only the scores are aggregated, no data leaves any lab.

Lab 1
FL participation: ACTIVE
Leave federation
pod data | fl config

Lab 2
FL participation: ACTIVE
Leave federation
pod data | fl config

R² climbing → a lab leaves & re-joins (live consent) → outsider blocked from raw data (401)

What it means



Surveillance → foresight

the Cattle Barometer can now forecast, not just report



Privacy by design

no raw data moves; the server is blocked from reading it



Welfare & fewer antibiotics

early warning supports timely, targeted action



Scales across Europe & species

add a POD + agent; extends to pig & poultry barometers

Conclusion

Ontology-driven federated learning predicts BRD one month ahead across 6 European laboratories and 4 countries — with accuracy comparable to centralized training, and without centralizing a single record.



The Livestock Health Ontology aligns heterogeneous feature spaces (FedAvg-ready).



GCN captures spatial structure; GRU captures temporal trend.



A viable path to privacy-preserving collaborative disease surveillance.



The data never leaves. The intelligence does.

Thank you

Saba Noor · saba.noor@ugent.be

DECIDE · EU Horizon 2020 · Grant Agreement No. 101000494

