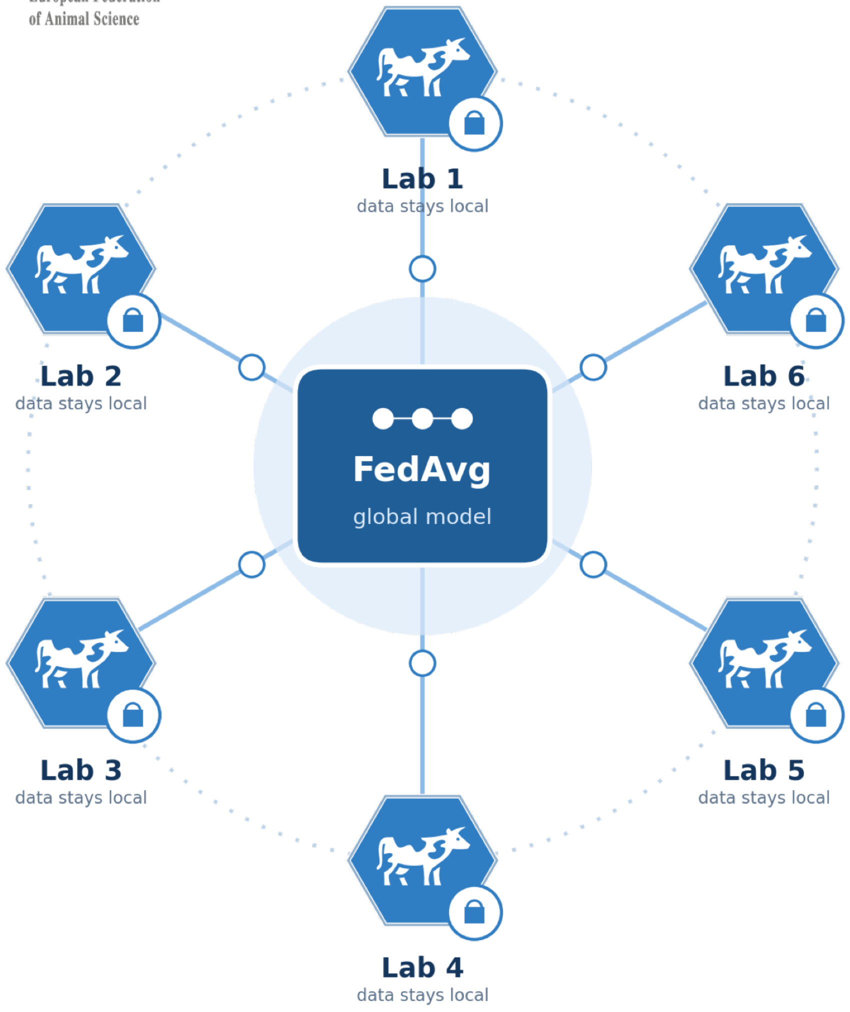
EAAP European Federation of Animal Science



Federated Learning for Bovine Respiratory Disease prediction Across European Veterinary laboratories

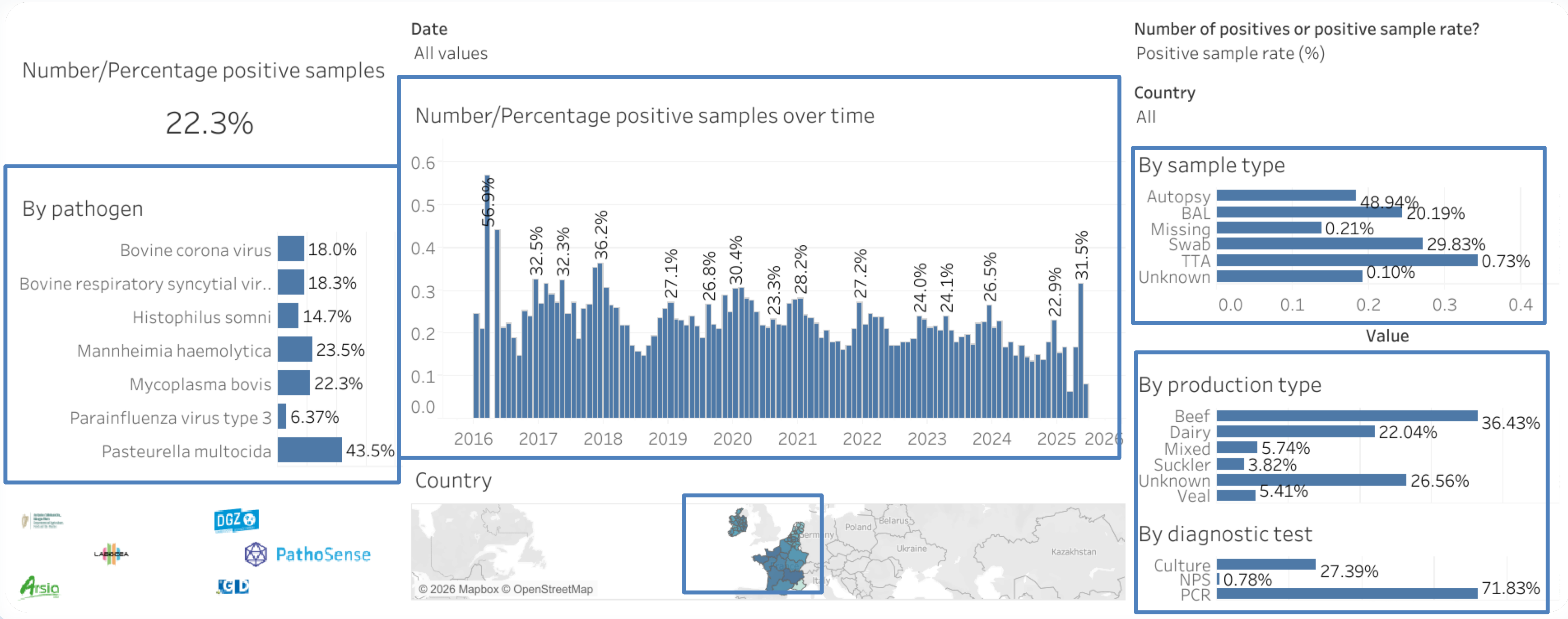
The data never leaves. The intelligence does.

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

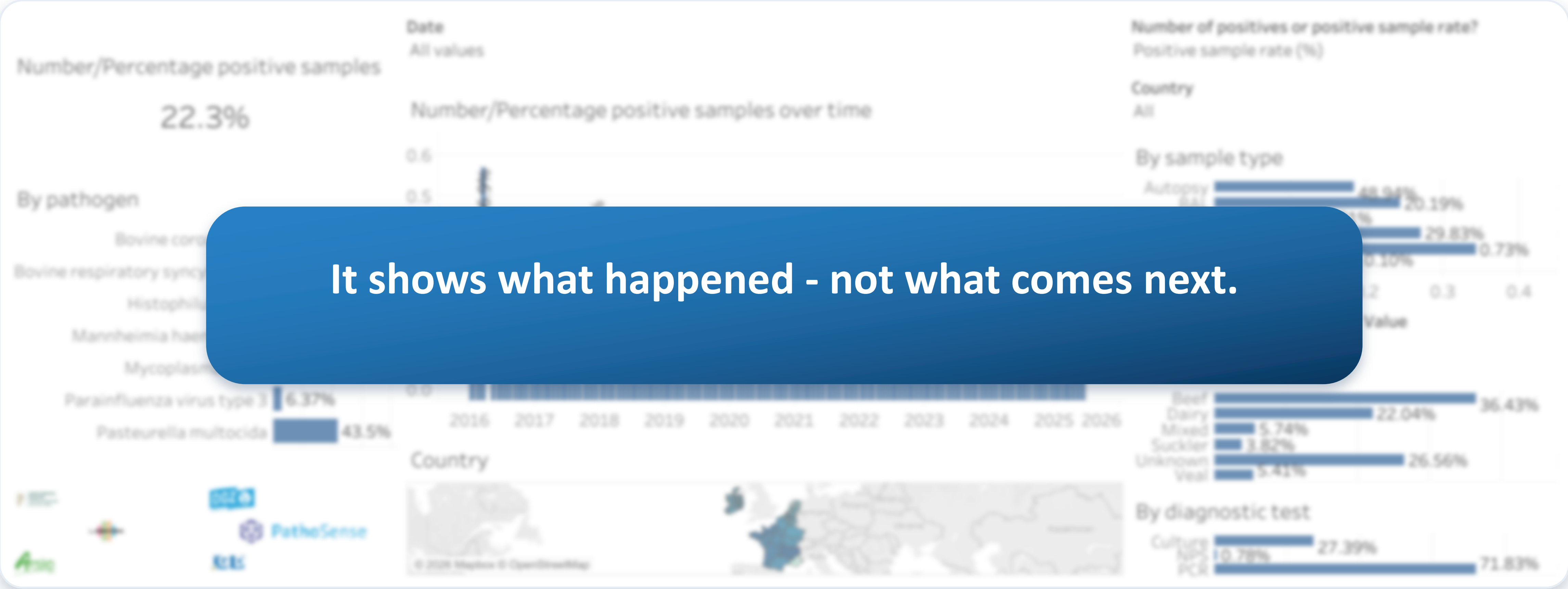
Scan to explore
the live tool



22.3% positive · 7 BRD pathogens · 4 countries (BE · NL · IE · FR) · 2016-2025

The Cattle Barometer today

Rich history - but it only looks backward



It shows what happened - not what comes next.

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

Raw CSV / Excel
per-lab files

→

Clean · anonymize · harmonize
per-lab preprocessing

→

Common semantic layer
map to LHO → RDF

→

Modeling set
one-month-ahead
shared baseline for 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)

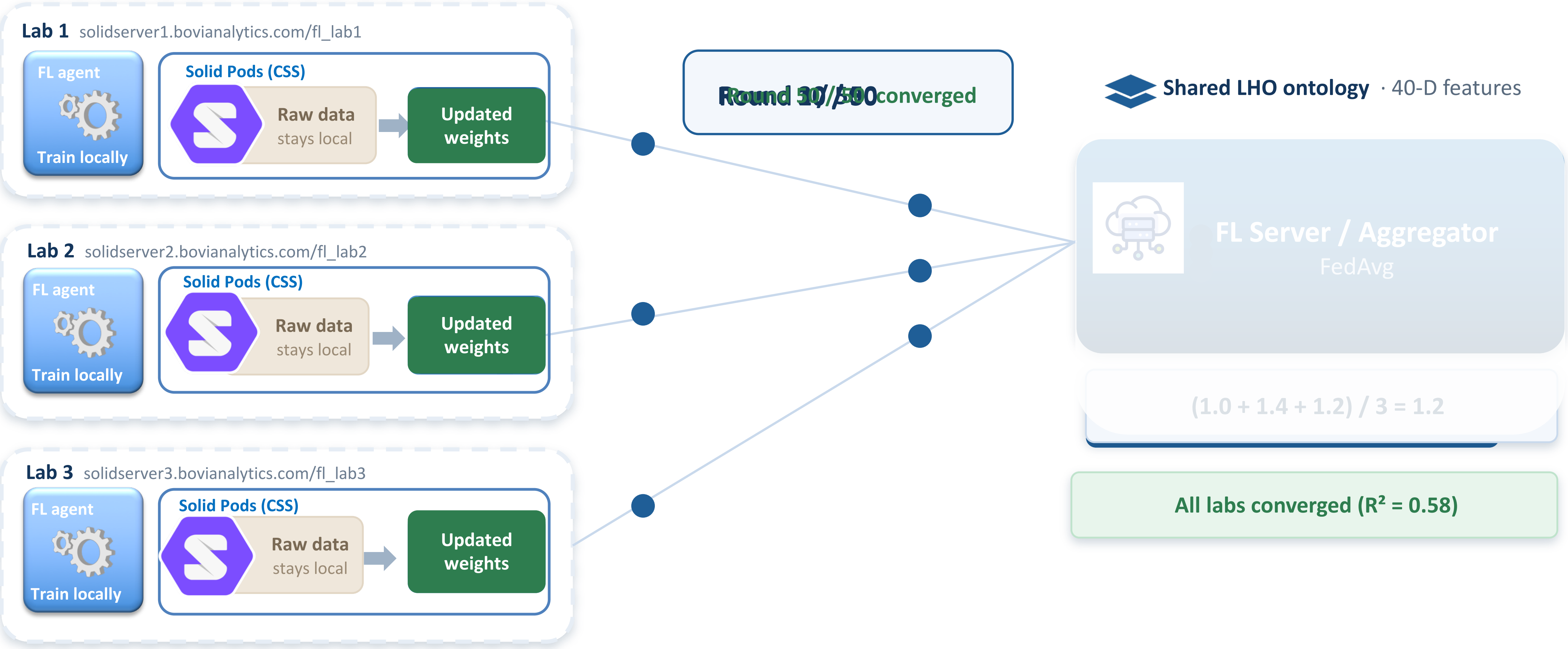
The model travels to each lab's Solid POD, trains locally — only the learned weights return. Raw data never leaves.

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.**

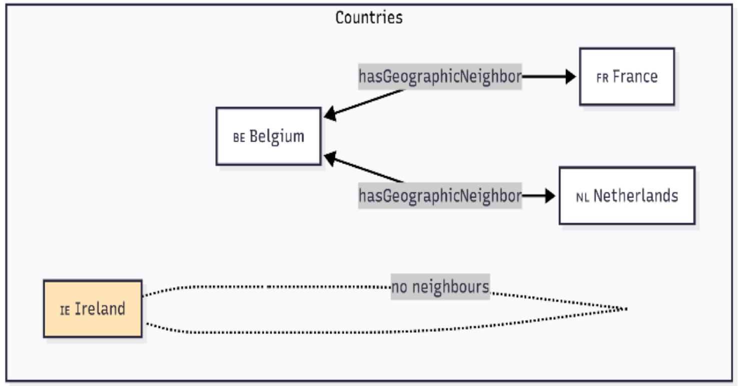
Federated learning in action

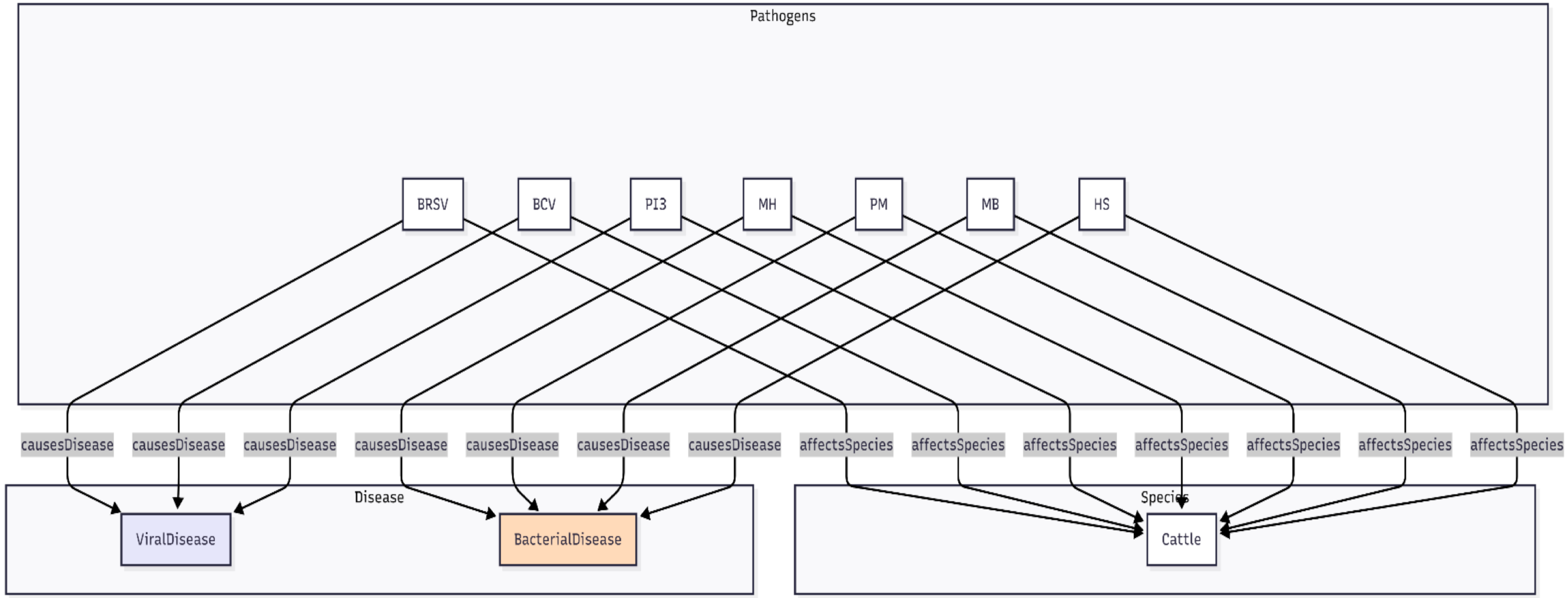
FedAvg over Solid PODs — weights travel to the server; raw data stays locked in each lab.



1 · Train locally 2 · ... 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

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

Scan to explore it live

Browse live - AgroPortal

Source code - GitHub

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)

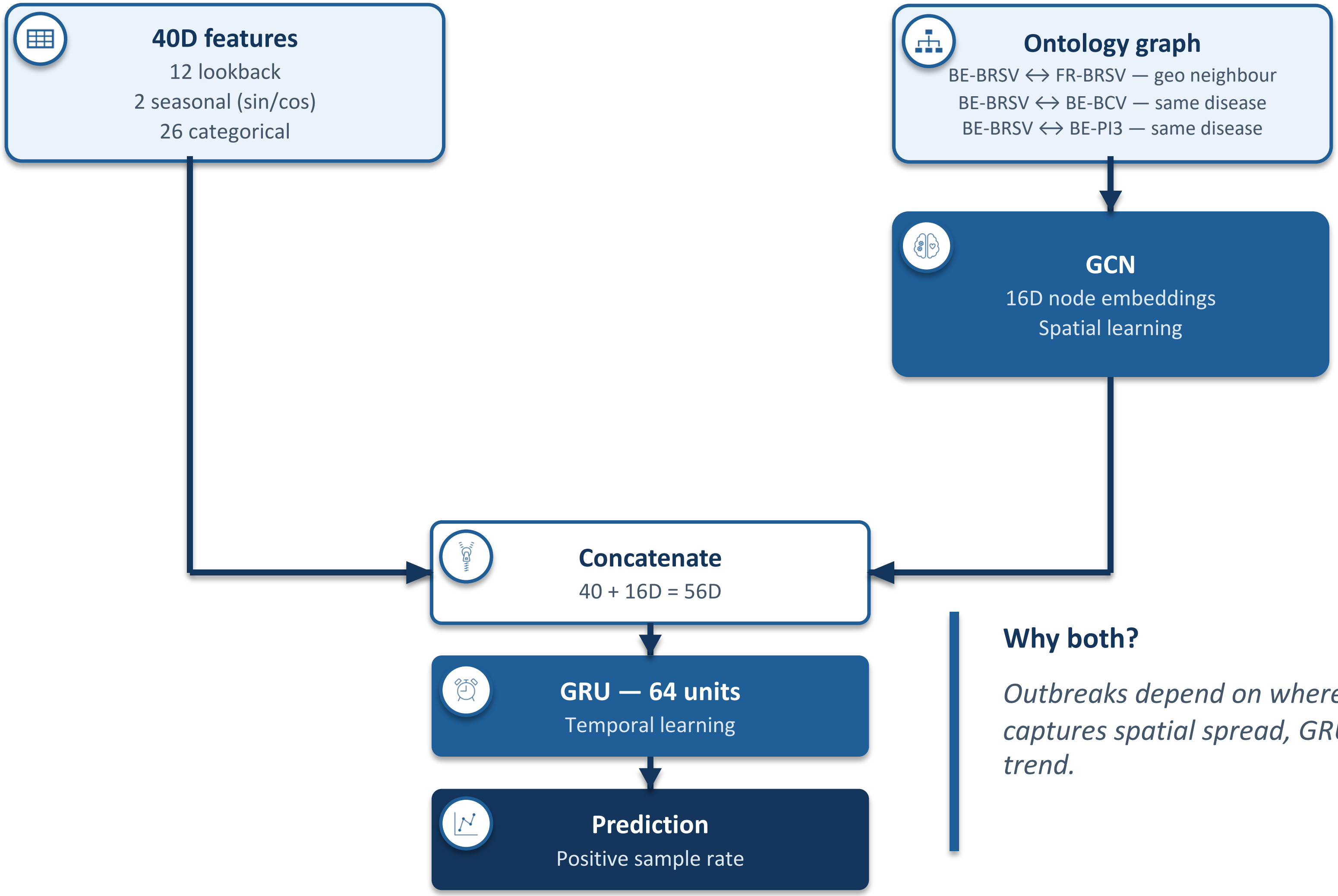
(zero for any it doesn't have)

Same dimensions → FedAvg works ✓

Queried via SPARQL — extract categories & relationships

The model: GCN + GRU

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



Why both?

Outbreaks depend on where and when — GCN captures spatial spread, GRU captures temporal trend.

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

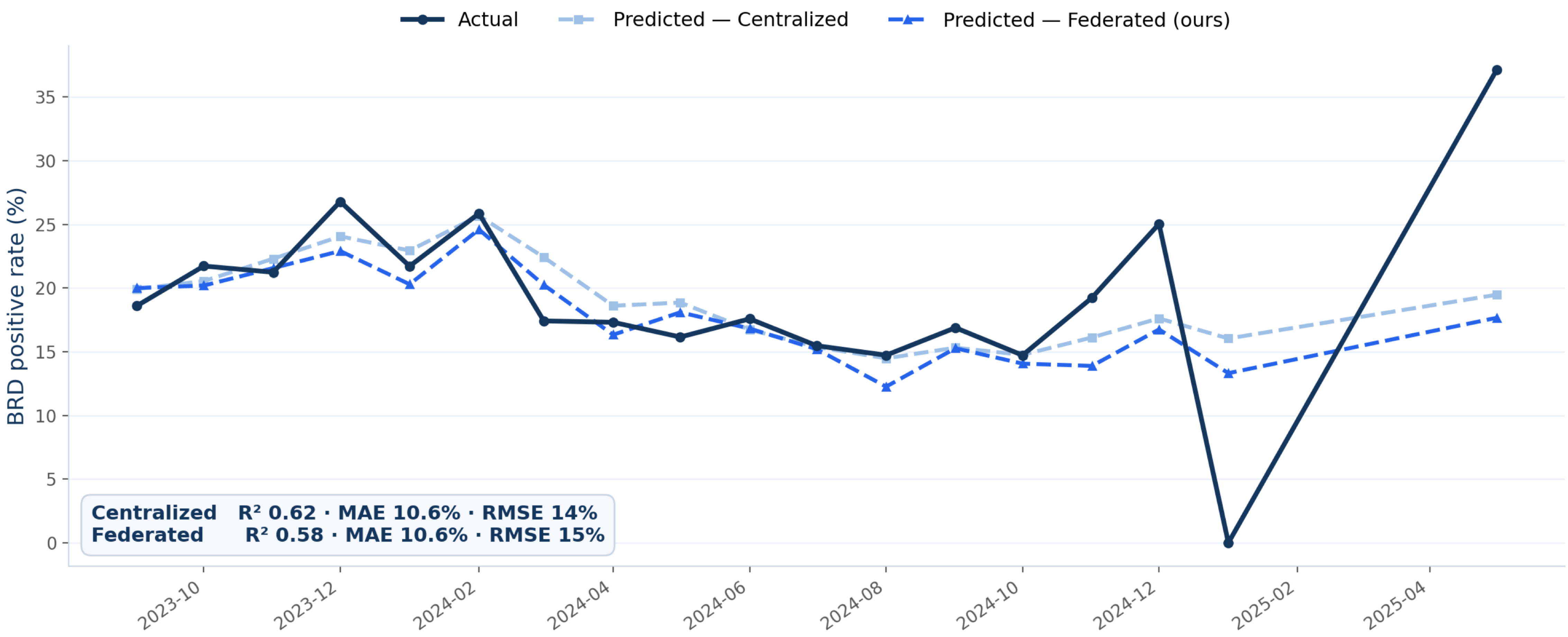
Common evaluation protocol across both modules

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



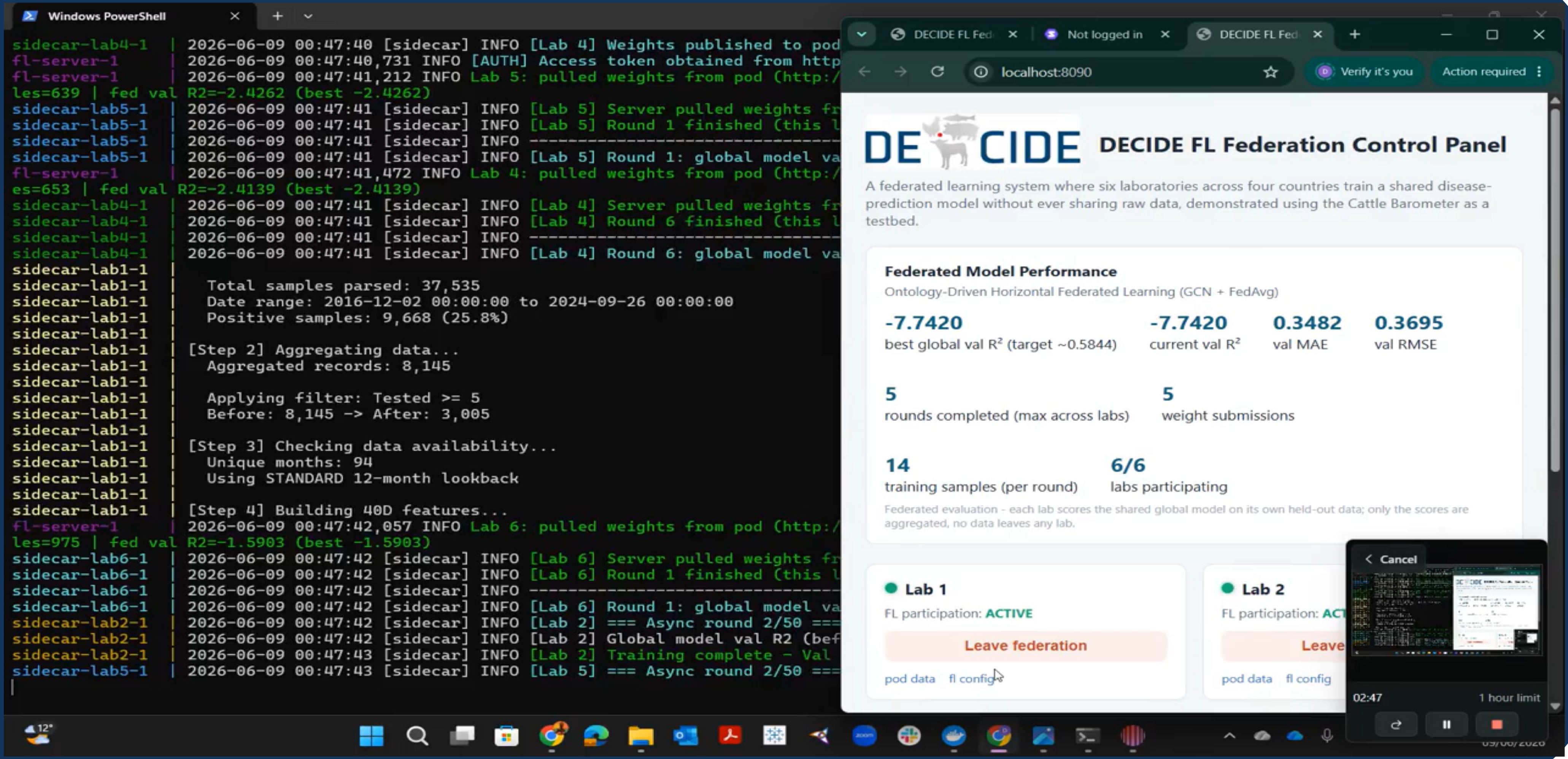
One-month-ahead BRD-positive rate · test period Sep 2023 – May 2025 · pooled over 4 countries · 7 pathogens · 6 labs

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



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