A review on implementational gaps and barriers regarding data quality and robustness for AI applications in livestock digital solutions

A. Lebreton, **C. Allain**, J. Niemi, M. Pastell, A. Stygar

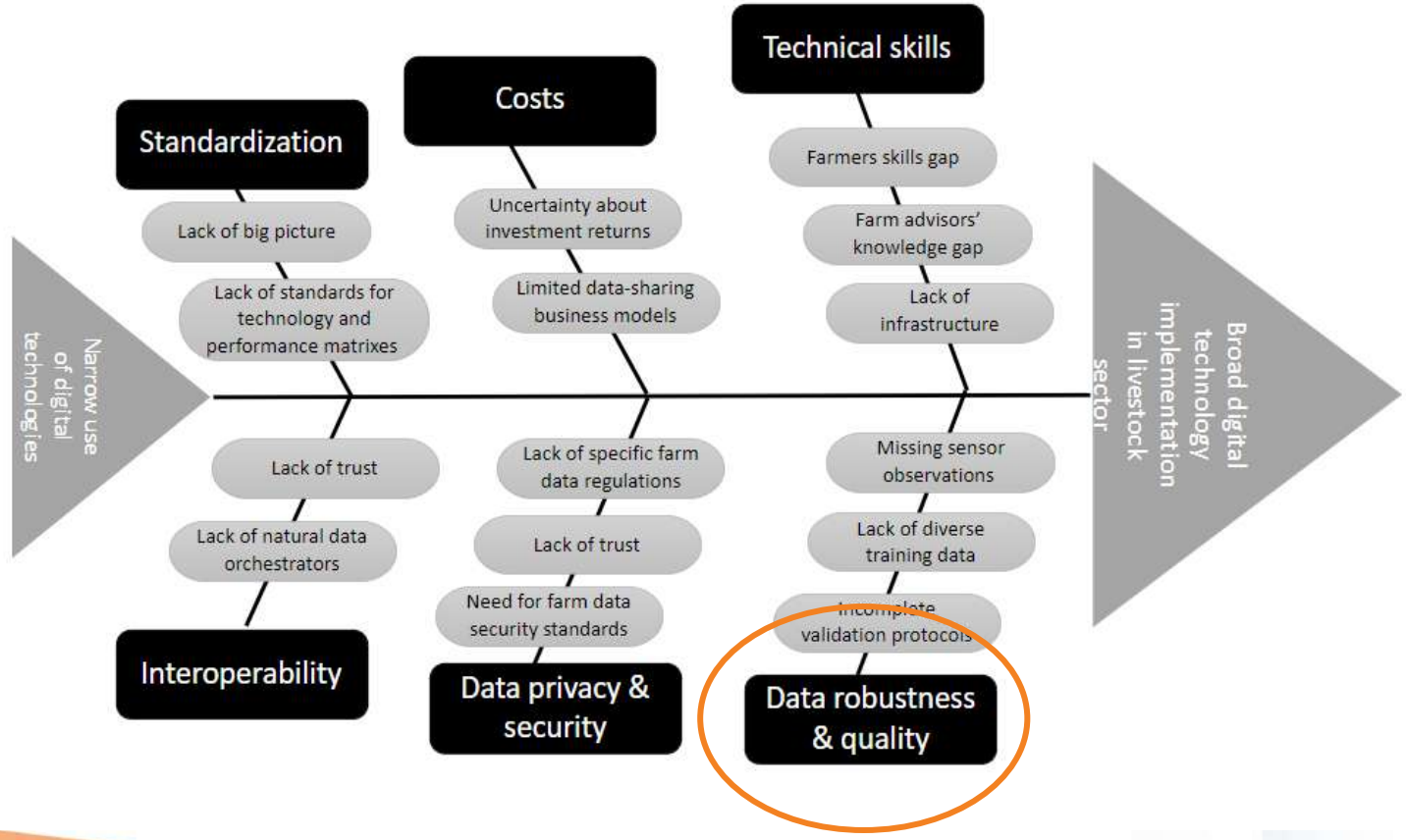
Clement.allain@idele.fr

Funded by the European Union

The Digi4live project

Digi4Lives's goal is to help **the European livestock sector** leverage **digital technologies** and data for the **benefit of farmers** and **food** and **technology companies** and to facilitate **public administration.**

Opinion paper on **gaps, barriers and solutions to digital technologies adoption** in the livestock sector



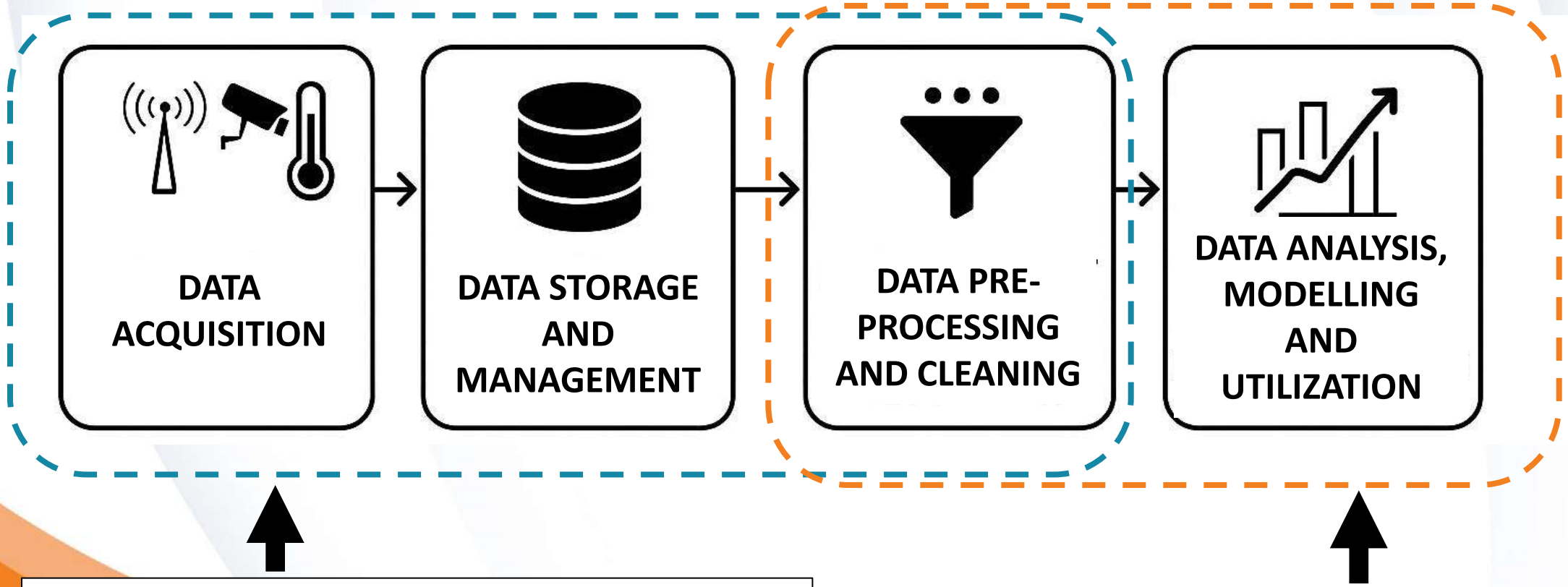
Funded by the European Union

The Critical Role of Data Quality and Robustness in the AI Era

Artificial intelligence is not magic. **It only knows what it has learned**

Funded by
the European Union

Data quality and robustness : what does it mean ?

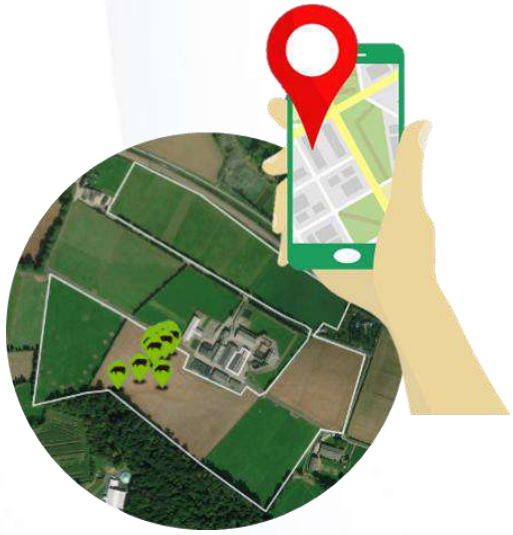


Quality = completeness, accuracy, consistency and reliability of data

Robustness = ability of data to stay useful and interpretable even in imperfect conditions

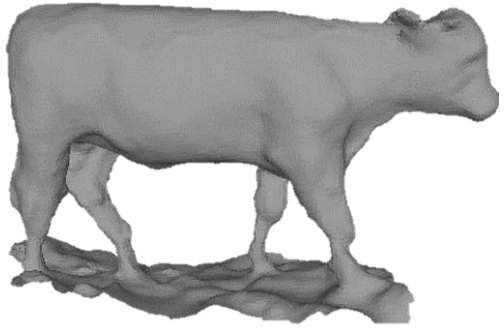
Funded by the European Union

Illustration based on 2 use cases



Grazing traceability with GPS solutions

PHENO3D



High throughput phenotyping based on 3D imaging

Funded by the European Union

Grazing traceability with GPS solutions

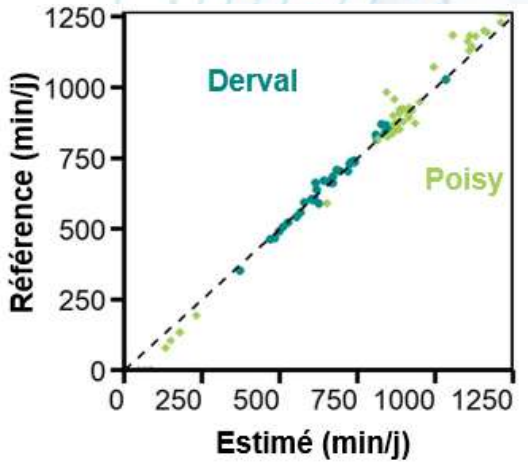
GPS collars

Cow location (LPWAN transmission)

Data preprocessing +
Paddock Detection models

Grazing time calculation models

Paddock grazed
Nb hours/day
Nb days grazed
…



Funded by
the European Union

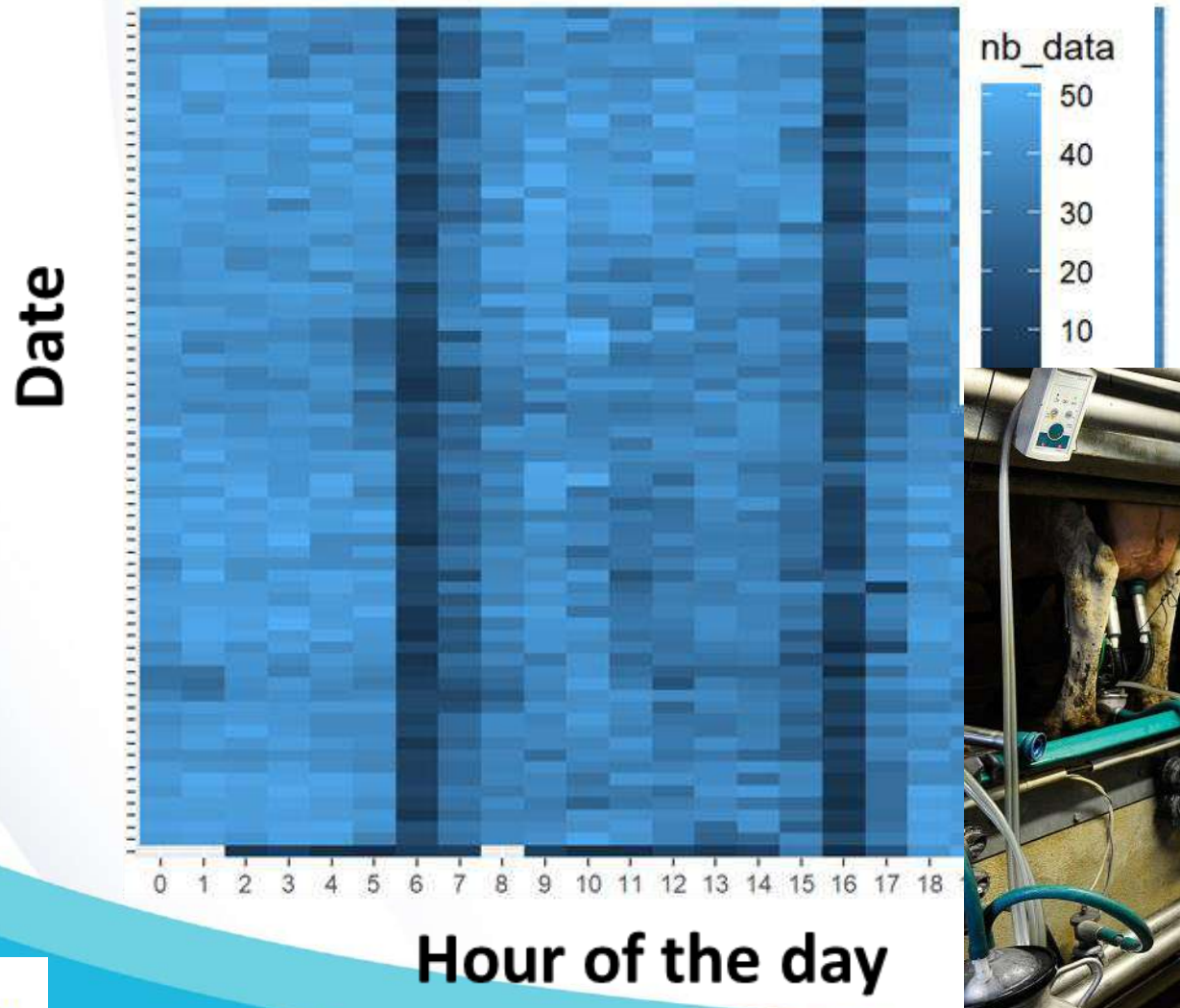
Grazing traceability with GPS solutions : Incomplete data

Data incompleteness

→ Effect of mountain/sea on Sigfox signals

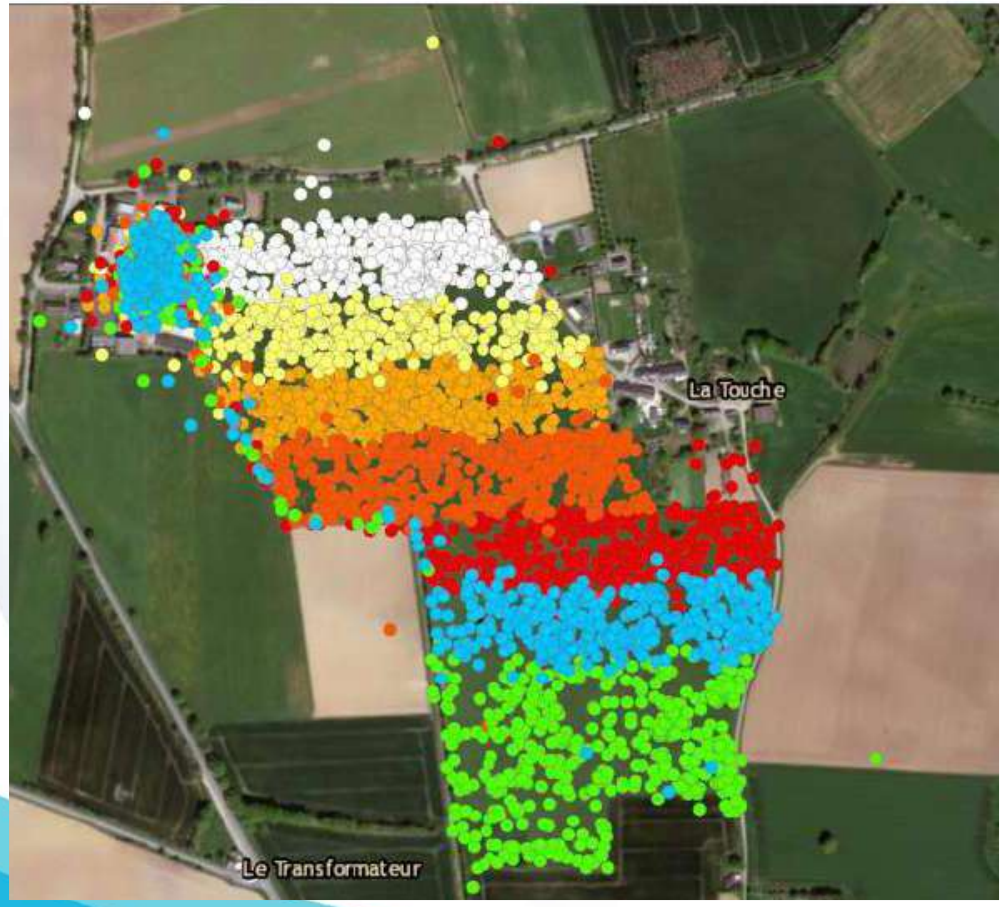
Funded by the European Union

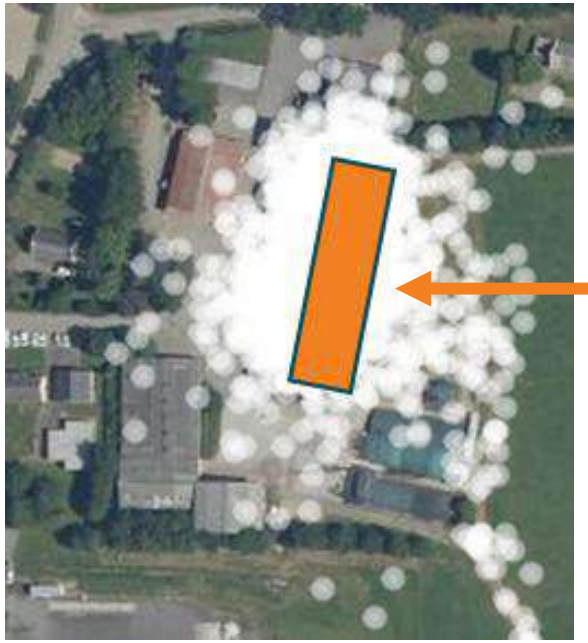
Grazing traceability with GPS solutions : Incomplete data



Funded by
the European Union

Grazing traceability with GPS solutions : Data inaccuracy





True barn position

Funded by the European Union

Grazing traceability with GPS solutions : solutions vs. barriers

Use advanced, reliable sensors, and adapted IOT networks ?

- GPS + sigfox
- 3 months disposable battery
- 250 €

- Multi-protocol (Satellite, 5G, LPWAN,…)
- Solar panels
- Buffer memory
- 1000 €

Too costly !

→ **Necessary trade-off** between **improving data quality and enhancing robustness**

Funded by the European Union

Grazing traceability with GPS solutions : solutions vs. barriers

Expert-driven preprocessing rules

Locations outside any polygons

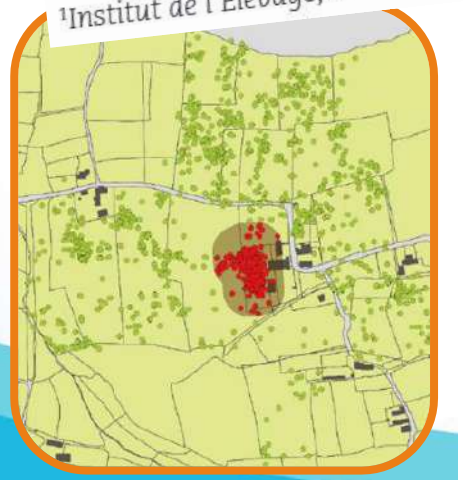
GNSS inaccuracy due t the barn effect

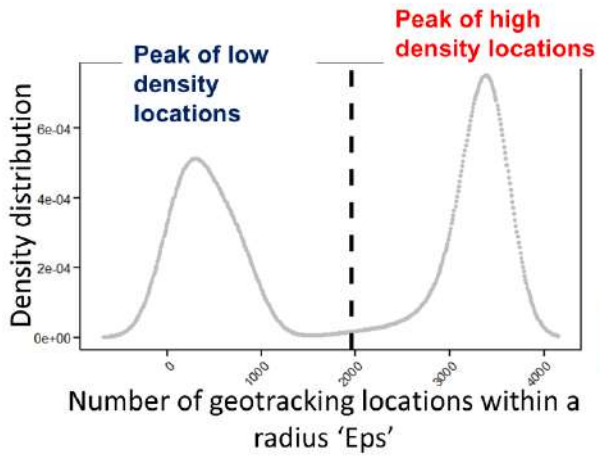
Kinetic corrections of the labels

PPBPP -> PPPPP

Are GPS sensors and density-based classification suitable to ensure the traceability of dairy cows on pastures? Part I: Development and validation on experimental farms

A. Lebreton[1], C. Allain[1], C. Charpentier[1], M. D'Introno[1], A. Fischer[1], W. Lonis[1], E. Nicolas[1], A. Philibert[1]
[1]Institut de l'Elevage, 149 Rue de Bercy, F-75595 Paris, France





« barn » locations

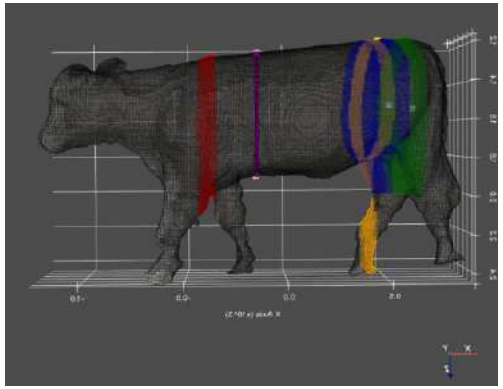
« pastures » locations

DBSCAN (Hahsler et al., 2019)

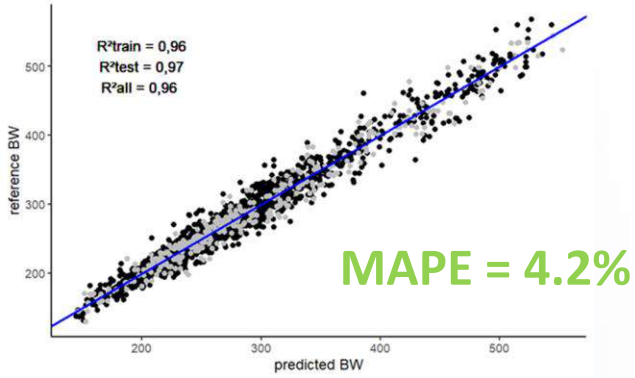
Funded by the European Union

3D imaging phenotyping : Phéno3D project

3D Scan



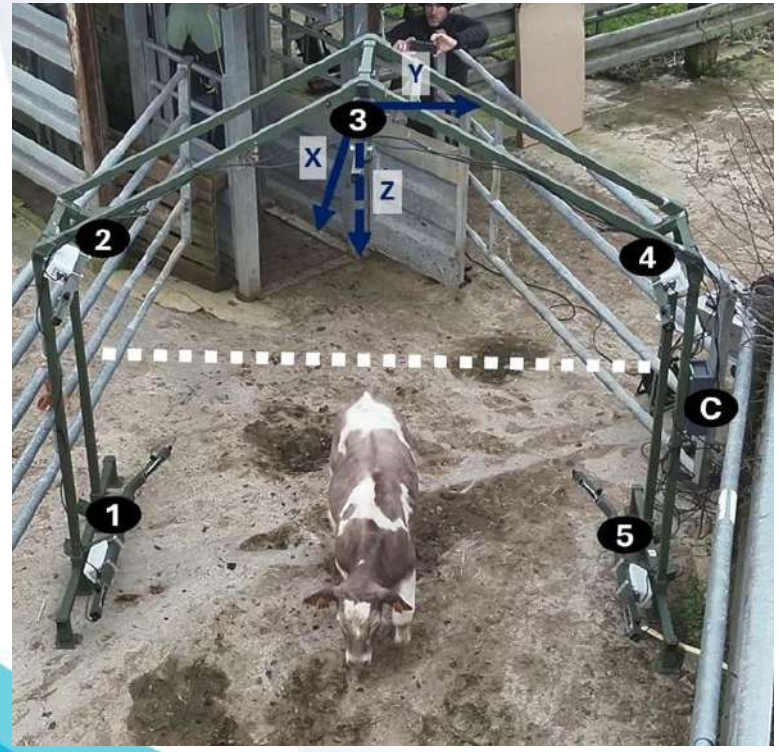
Data preprocessing
(See Do et al. (ECPLF 2024))



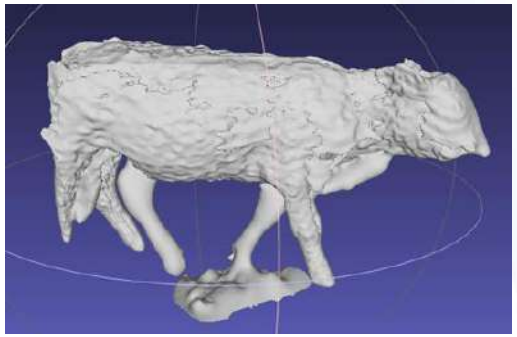
Body weight Prediction
(See Dechaux et al. (ECPLF 2024))

Funded by the European Union

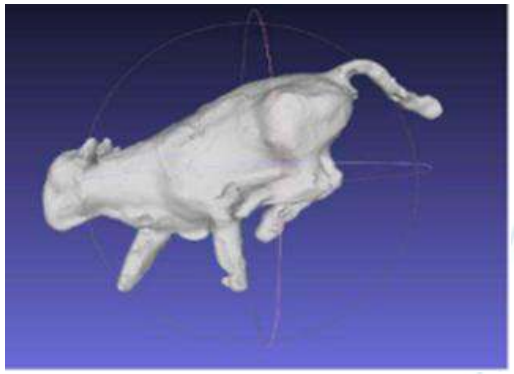
3D imaging phenotyping : how to deal with noise ?



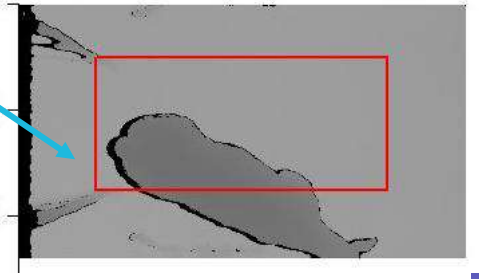
Sun

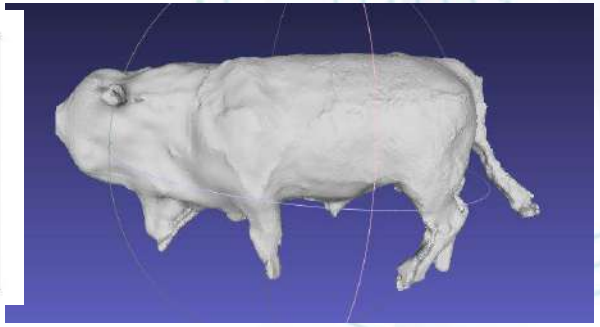


Jump



Not centered



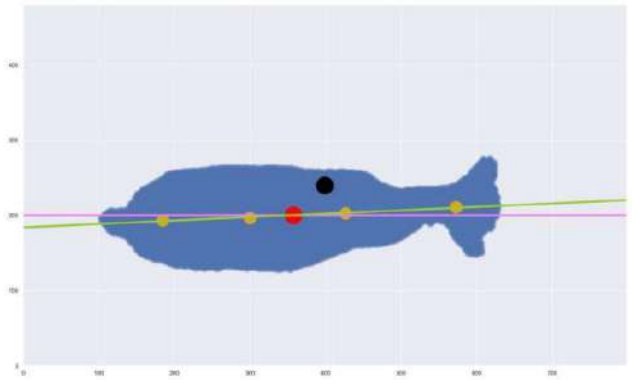


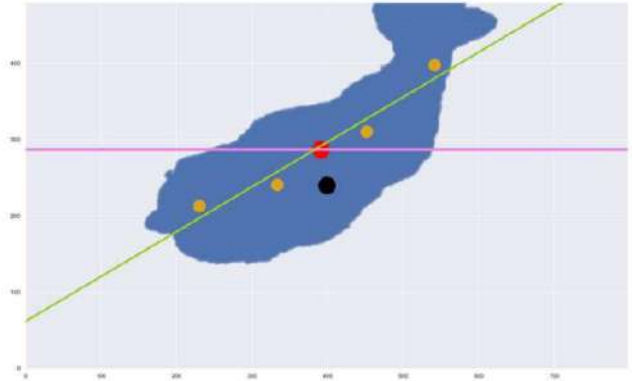
Funded by the European Union

Solutions at the data collection stage

- Improve animal handling

- Develop automatic data quality assesment

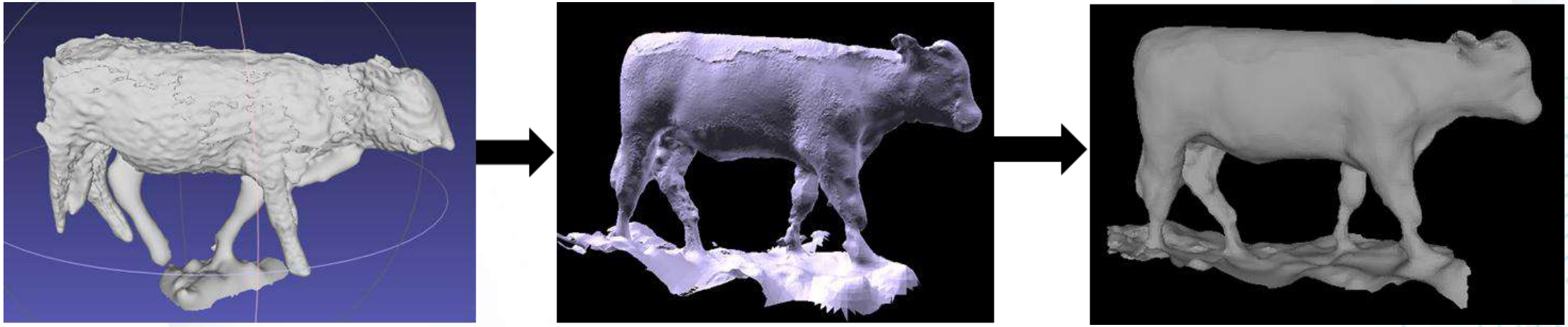




Funded by
the European Union

Solutions at the data pre-processing stage

- Improve filtering, smoothing, calibration techniques



Funded by
the European Union

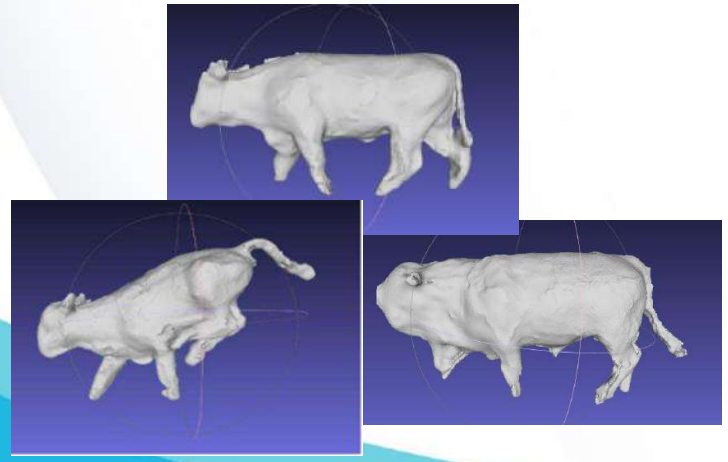
Solutions at the model development stage

Train robust machine learning models with noisy data !

All data (2079 images)

Body weight prediction

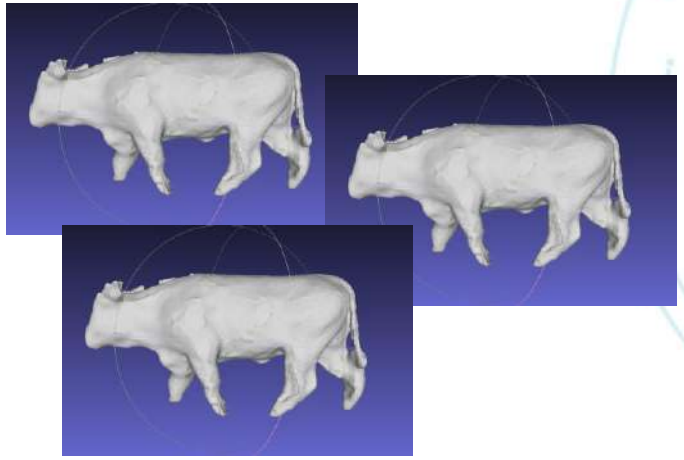
(MAPE = 4.2%)



All data - 25% of worse images

Body weight prediction

(MAPE = 3.9%)



Funded by the European Union

Conclusion : Please mind the gaps !

- **Quality**
  - ➢ Embedding it in the validation and impact assessment of digital tools in research and practice.

- **Robustness**
  - ➢ As imperfect data becomes the norm, the robustness of AI solutions must rise. Stronger, more resilient Proofs of Concept are essential to ensure real-world reliability and scalable deployment.

Stay tuned – more insights coming soon from the **Digi4Live** project!

Funded by the European Union

Summary of gaps, barriers and solutions

Gaps	Barriers	Solutions
Incomplete data due to sensor failures, connectivity loss, or low battery life.[36]	High costs of advanced, reliable sensors, and adapted IOT networks. [36]	Enhance energy storing, connectivity, buffer memory; redundancy & interoperability for cross-system validation [36]
Lack of best practice guidance in dealing with noise.	High costs of advanced, reliable sensors; lack of proper calibration and maintenance.	Use advanced data preprocessing techniques to clean and correct anomalies.
Precision and accuracy - poorly calibrated sensors or sensor drift. [37]	Data analysis strategies effect on data accuracy and precision (statistical vs. machine/deep learning) [39]	Regular sensor recalibration and maintenance; on-farm validation for robustness in real conditions. [39]
Temporal synchronization and interoperability – Inconsistencies [37]	Difficulty in aligning multiple sensor outputs; challenges in ensuring seamless interoperability. [37]	Standardized protocols for synchronization. Encouraging interoperability between systems. [37]
AI Data Robustness -Limited diversity in training datasets, bias and overfitting. [40]	High costs and difficulty of collecting diverse, high-quality datasets; over-reliance on quantity rather than quality. [40]	Follow "Ethics guidelines for trustworthy AI"[35]. Foster shared datasets, incorporate data of various regions and conditions; prioritize variance & quality [40]
Tool Validation - Incomplete protocols for completeness or repeatability.	Limited on-farm validation and adoption of comprehensive validation protocols.	On-farm validation under variable conditions; Promoting broader validation standards.
Stakeholder Engagement – How data quality impacts farms and stakeholders.	Lack of training and awareness about digital tools and their potential consequences.	Training of farmers and stakeholders; promote clear communication and equipment maintenance.

Funded by the European Union

Thank you!

digi4live.eu

Clément Allain
French Livestock Institute (Idele)
clement.allain@idele.fr

Funded by the European Union

Digi4Live Digi4Live Digi4Live

@Digi4Live digi4live Digi4Live