Plant & Food Research
Rangahau Ahumāra Kai

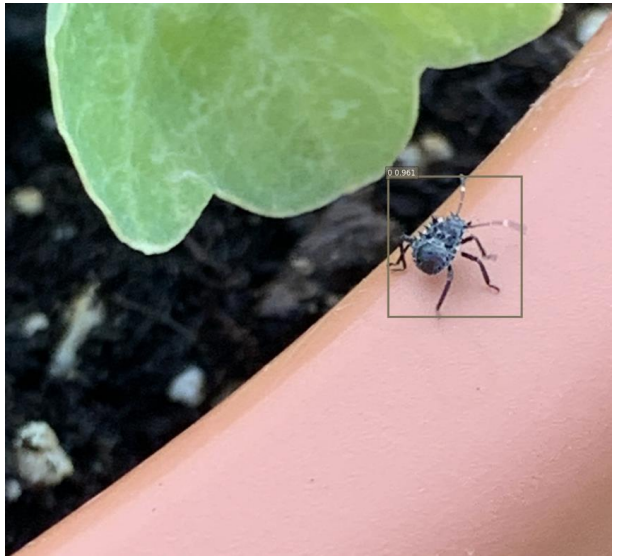
Building an MLOps stack for rapid delivery of reproducible computer vision projects

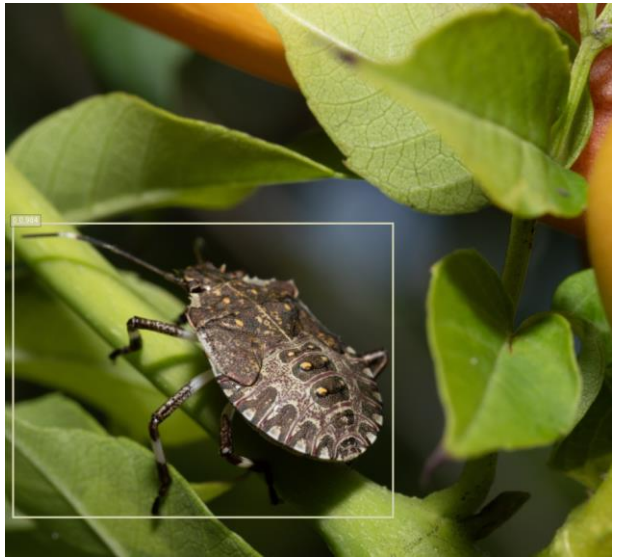
Daniel Bentall

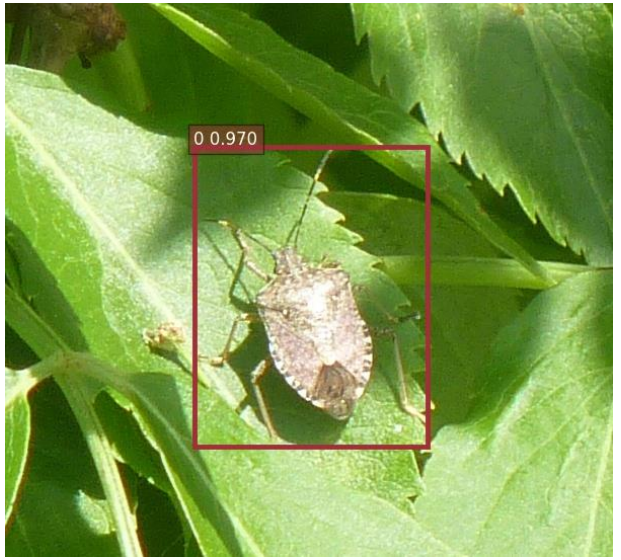
The New Zealand Institute for Plant and Food Research Limited

Object detection: brown marmorated stink bug

- Object detection labels
  - ImageNet: bugs, beetles and flies
- Classification labels
  - iNaturalist: stink bugs
  - PFR: stink bugs
- Train insect detector
- Inference on classification labels
- Train BMSB detector





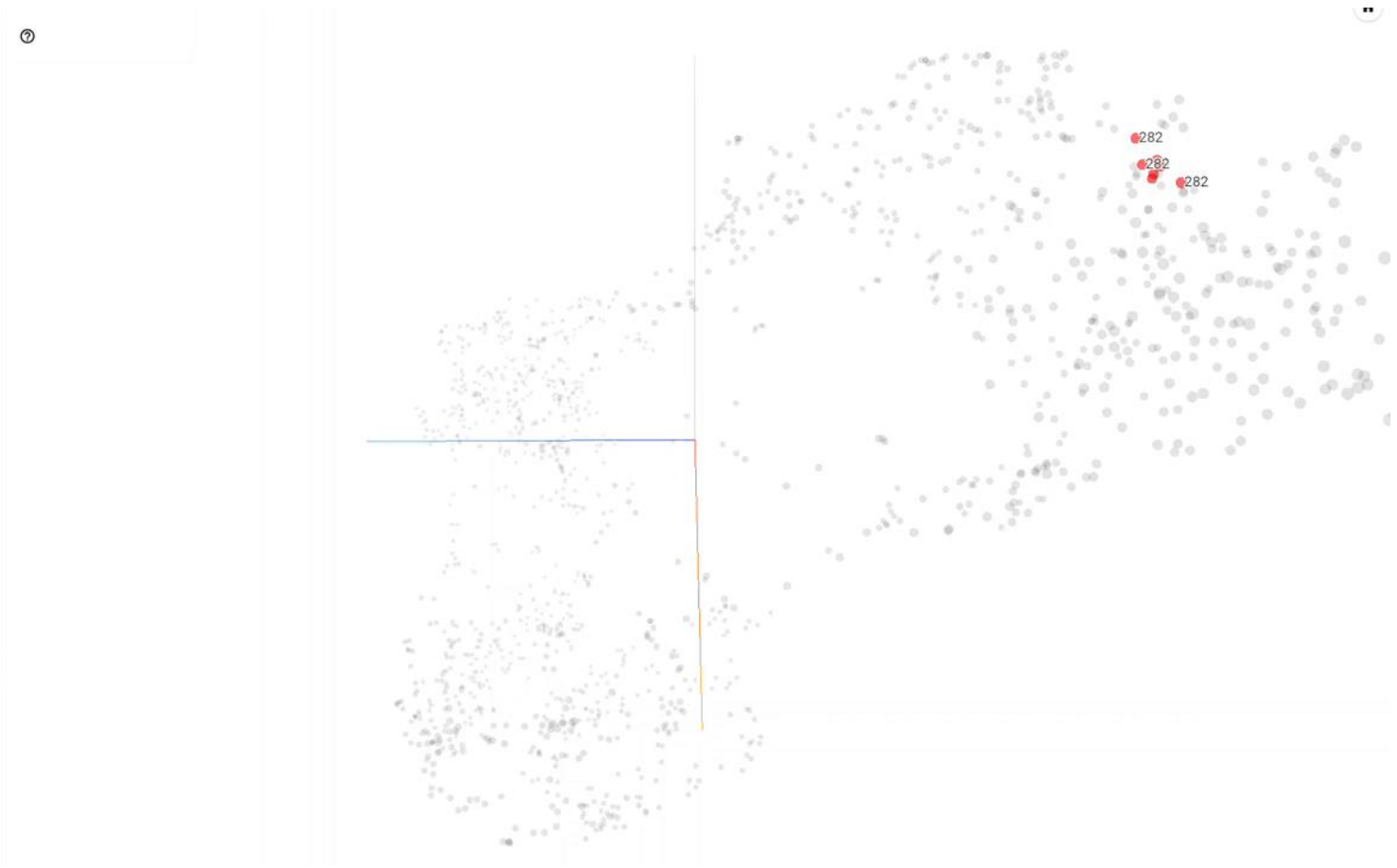


https://www.image-net.org/

https://www.inaturalist.org/

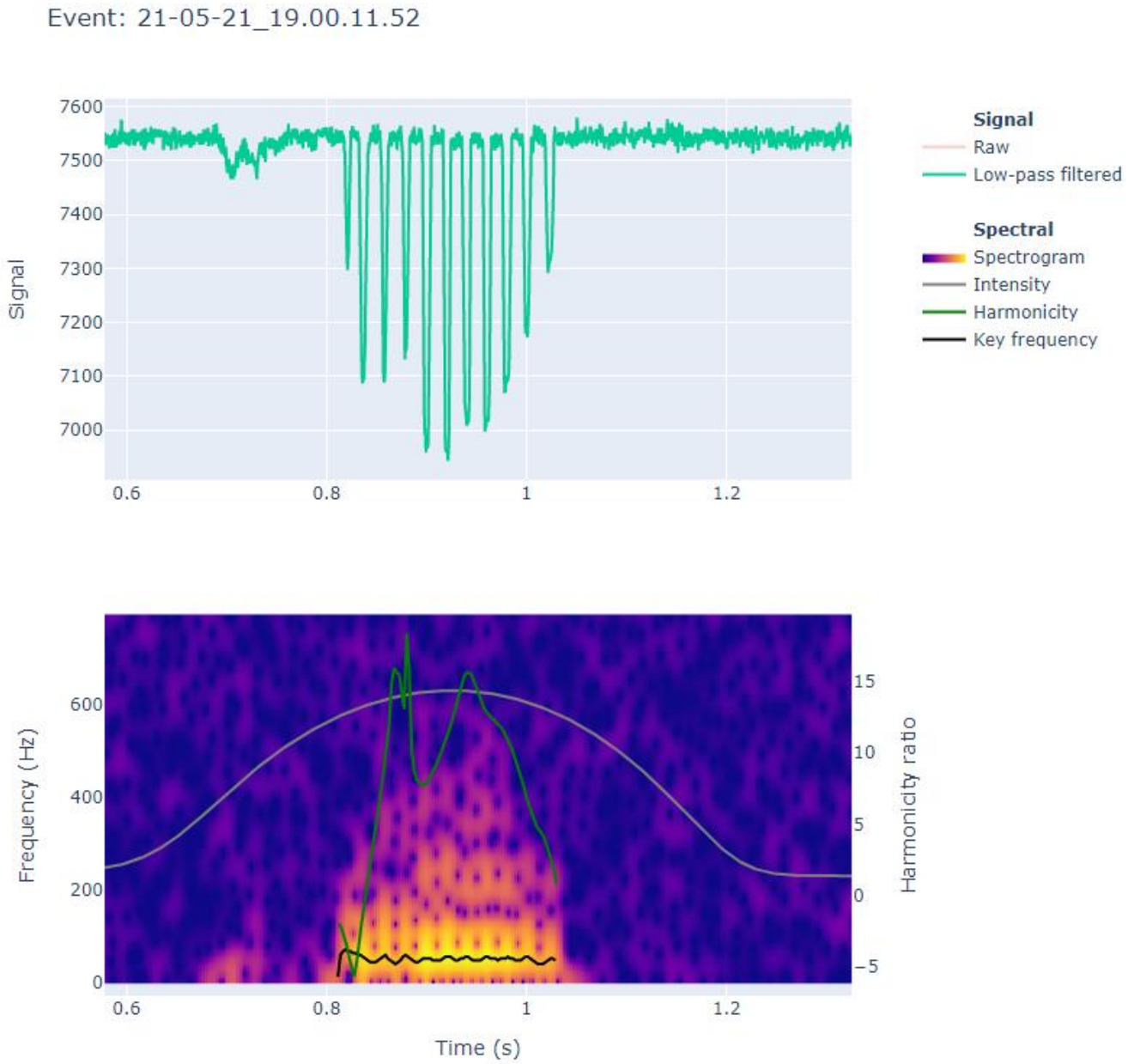
Identification by embedding

- Many similar classes (355), few labels per class (~3)
- Apply to new classes
- Embed model
- Learns distinguishing features
- Classify
- Cluster



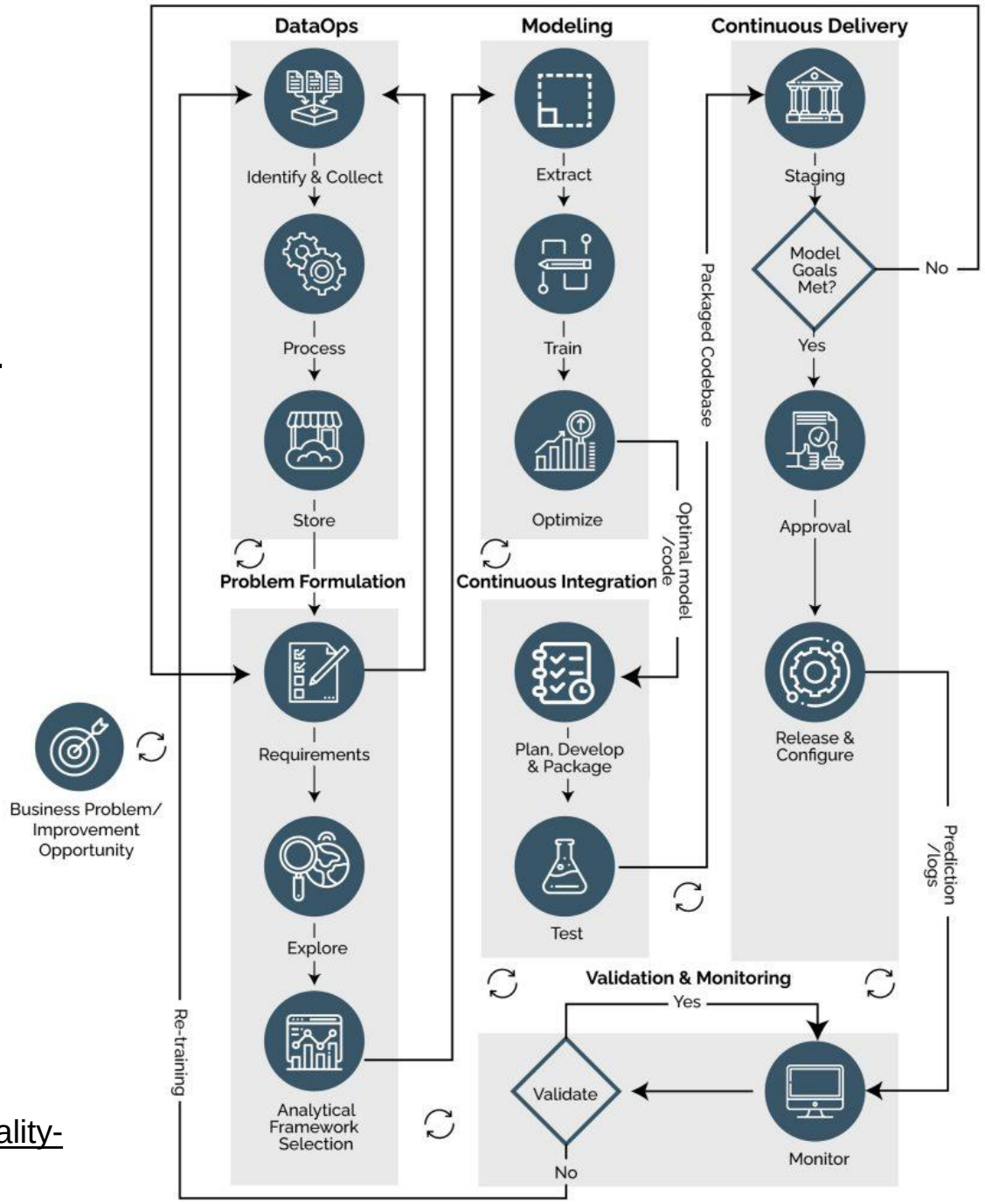
Spectral analysis for insect species classification

- Insect wingbeat signal
- Spectrogram
- Engineer features
- Or analyse as image
- RNN or transformer?



What is MLOps?

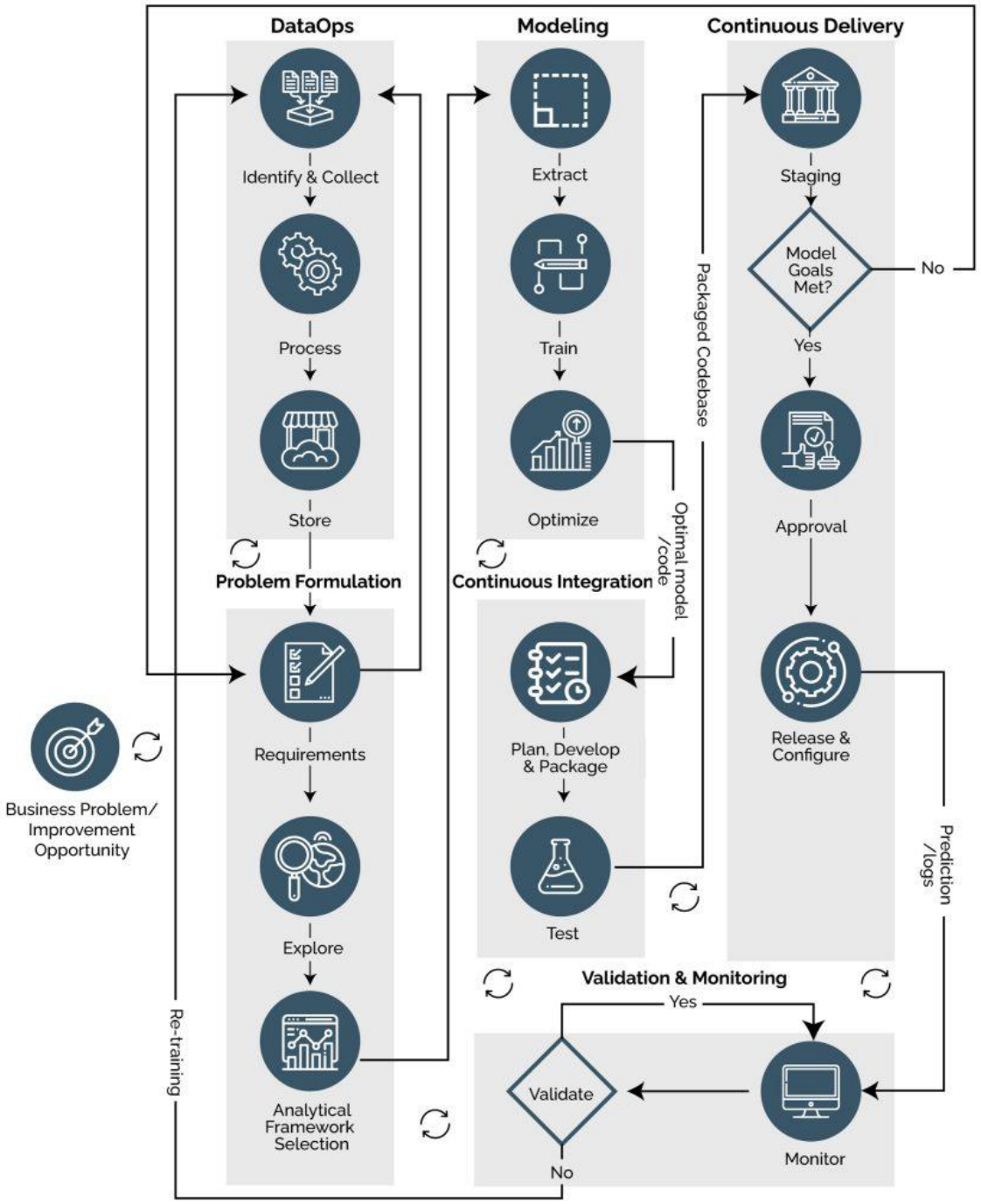
“MLOps is a set of practices that [combines] Machine Learning, DevOps, and Data Engineering practices for a reliable and data-centric approach to Machine Learning systems during production.”



https://radiant.digital/the-fundamentals-of-mlops-the-enabler-of-quality-outcomes-in-production-environments/

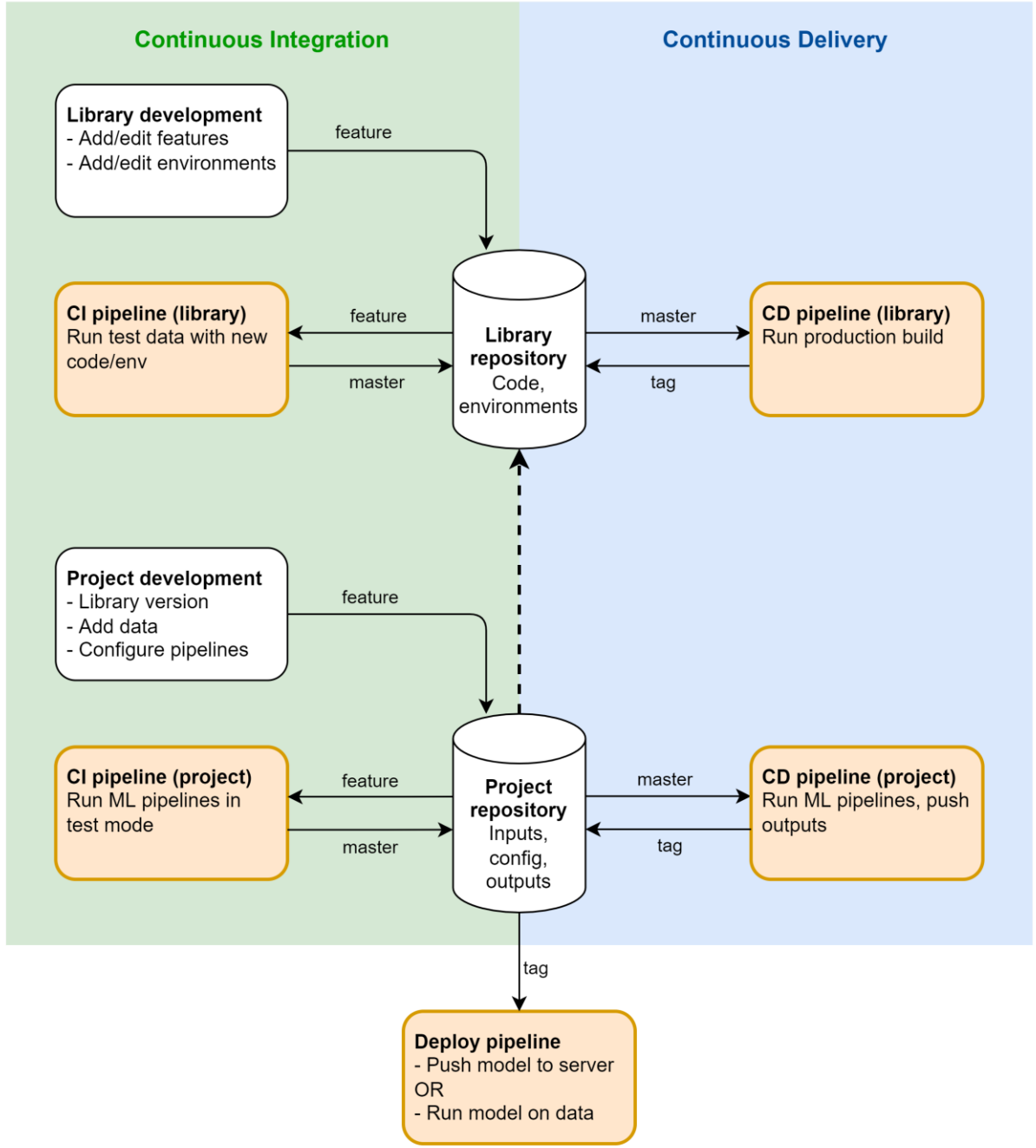
Our MLOps process and stack

- Project templating: Cookiecutter
- Image labelling: Supervisely
- Data storage: DVC (on git)
- Data exploration: Jupyter
- ML pipelines: Nextflow
- CI/CD pipelines: Jenkins
- Model serving: Multi-Model Server



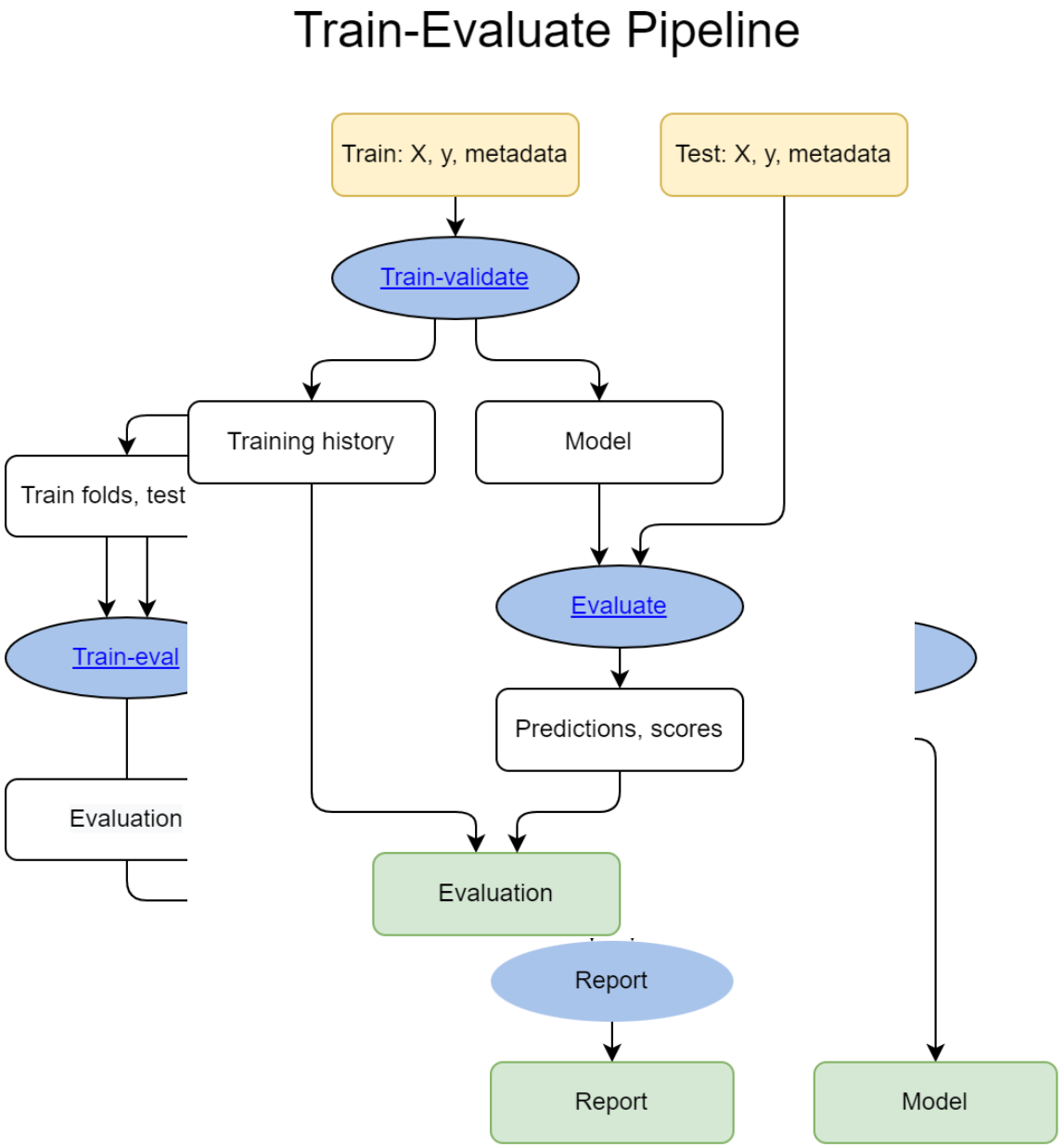
Development process

- Library and project repos
- Develop independently or together
- Development on feature branch
- CI tests and merges to master
- CD tags and produces models
- Deploy pipeline



ML Pipelines

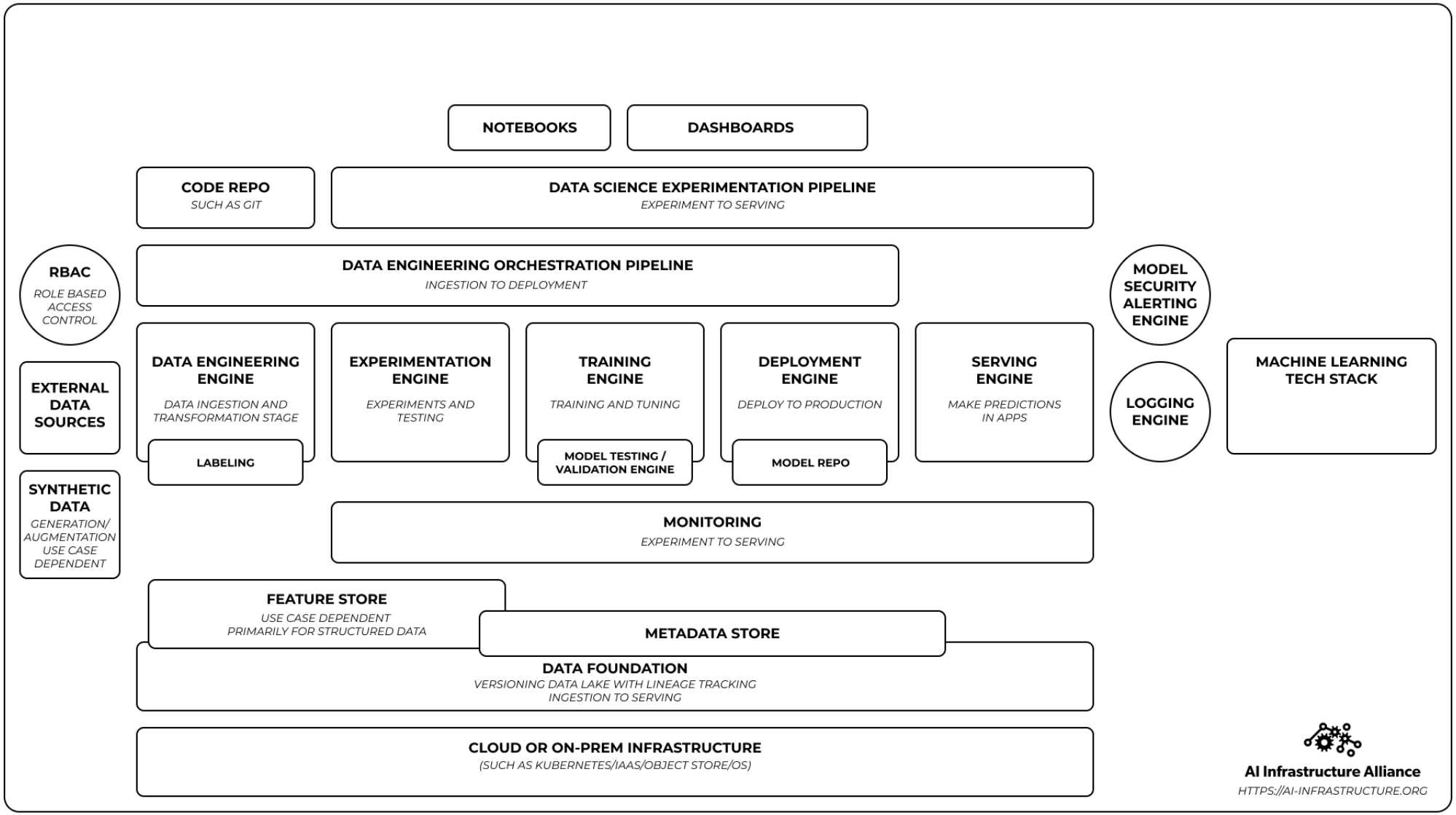
- Nextflow with DSL2
- HPC
- Environments
- Modular, nestable and extensible
- Configure at each level



Gaps and Future Work

- Production monitoring/validation
- Active learning
- Auto ML
- Explainable AI

MLOps Stack



https://opendatascience.com/the-rapid-evolution-of-the-canonical-stack-for-machine-learning/

Acknowledgements

- Harris Lin
- Sam Birch
- Krishna Babu
- Eric Burgueño
- Gonzalo Avila
- Flore Mas
- Taylor Welsh

Disclaimer

Presentation for:
eResearch 2022 conference, online, 09/02/22 – 11/02/22

Publication data:
Bentall D. February 2022. Building an MLOps stack for rapid delivery of reproducible computer vision projects. A Plant & Food Research PowerPoint presentation, SPTS No. 22073.

Presentation prepared by:
Daniel Bentall
Data Scientist, Data Applications
February 2022

Presentation approved by:
Linley Jesson
Science Group Leader, Data Science
February 2022

For more information contact:
Daniel Bentall
DDI: +64 3 325 9688
Daniel.Bentall@plantandfood.co.nz

Plant & Food Research
Rangahau Ahumāra Kai

Thank you

Daniel.Bentall@plantandfood.co.nz

plantandfood.co.nz