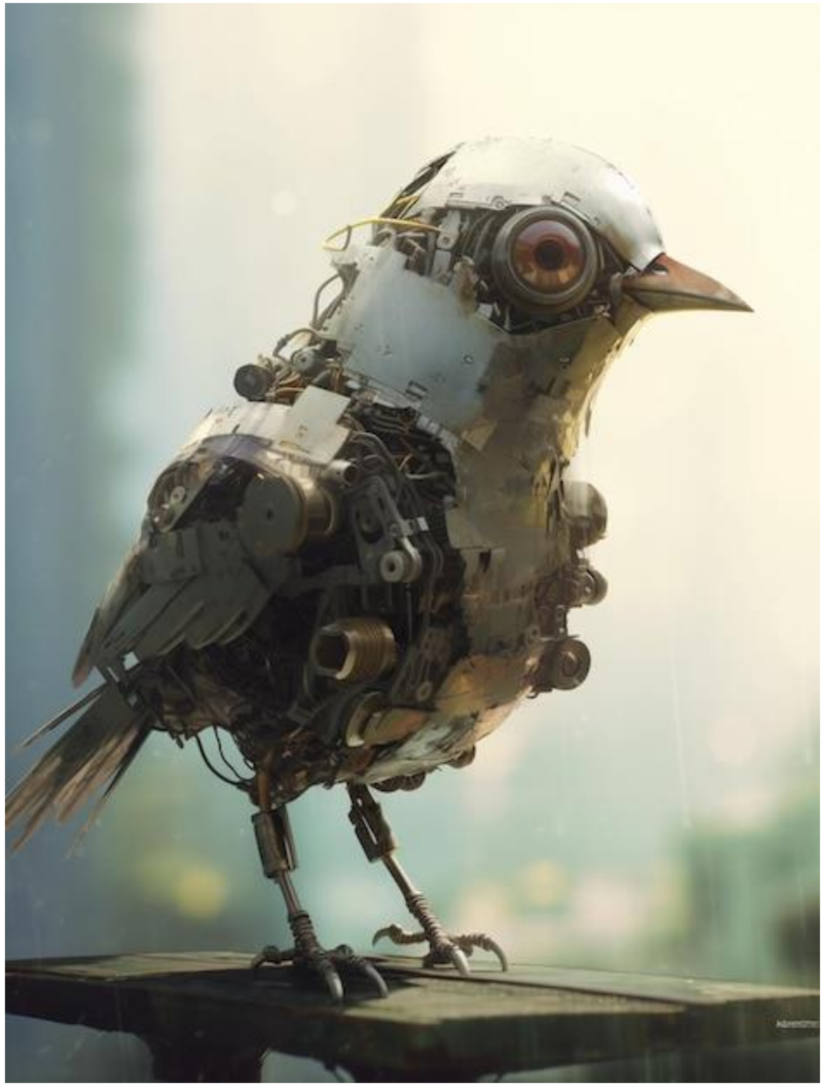
USE CASE 2 WEBINAR

S. JETSCHNY, M. VITTEK, R. KNAPEN, M. RAEISSI, G. SALVINI

16 APRIL 2024



CAN ARTIFICIAL INTELLIGENCE SUPPORT BIRD CONSERVATION?

FAIRiCUBE

WAGENINGEN UNIVERSITY & RESEARCH

AGENDA

10:00-10:15	Introduction to FAIRiCUBE	Stefan Jetschny
10:15-10:45	Use Case 2 (UC2) in a nutshell • Earth Observation for agriculture and biodiversity • Data cubes, AI and Causal ML • UC2 framework	Marian Vittek, Rob Knapen, Masoume Raeissi
10:45-11:15	Interactive session on preliminary results	Marian Vittek, Masoume Raeissi, Giulia Salvini
11:15-11:45	Interactive session on AI for biodiversity • How can AI support bird conservation? • Foreseen applications of UC2 outcomes • What are the benefits and limitations?	Rob Knapen, Giulia Salvini
11:45-12:00	Last announcements and closure	Giulia Salvini

FAIRICUBE – FACTS & OBJECTIVES

- Horizon Europe project, Research & Innovation Action
- Duration: 07/2022 – 06/2025
- High level objectives:
  - Establish **FAIRiCUBE Hub** – an integrated platform for data ingestion, analysis and ML application of FAIR spatial **earth observation data**
  - Demonstrate FAIRiCUBE Hub by running **5 use cases** addressing EU green deal actions (climate change, circular economy, biodiversity,..)
- FAIRiCUBE Hub
  - Harmonized data space,
  - Connect all the pieces required for a data science pipeline into a user-friendly framework,
  - FAIR (Findable, Accessible, Interoperable, Reusable),
  - TRUSTable (Transparency, Responsibility, User focus, Sustainability, and Technology),
  - Specific contribution to the Green Deal Data Space (GDDS).
- ***Deliver the power of data cubes and ML to decision makers and data scientists.***

CONSORTIUM

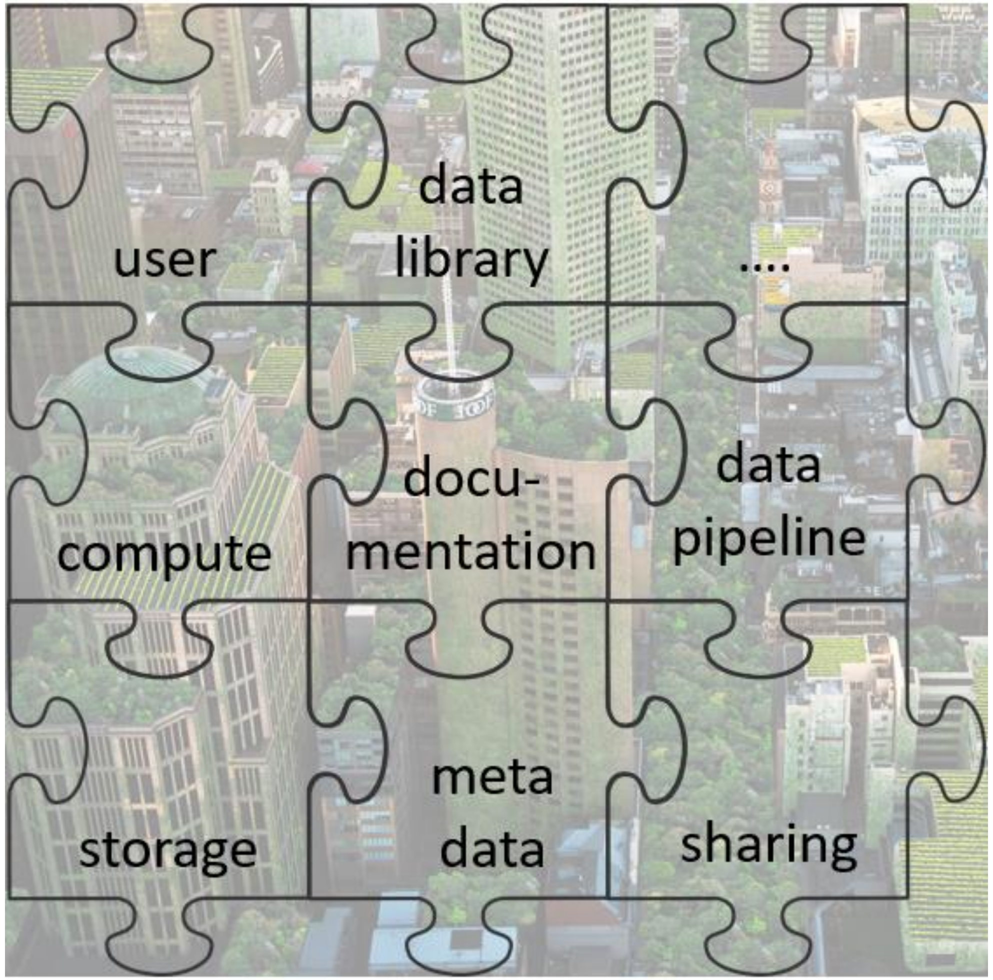
- Research institutes
  - NIL - NILU climate and environmental research institute, Norway
  - WUR - Wageningen university and research, Netherlands
  - NHM - Museum of Natural History Vienna, Austria
- Environmental SME's
  - S4E - space4environment, Luxembourg
  - 4SF - 4sfera, Spain
  - EPS - Epsilon, Italia
- Infrastructure service providers
  - EOX - EOX, Austria
  - CUB - Constructor University Bremen, Germany,
    supported by rasdaman GmbH

use case owner, geospatial data specialist, infrastructure specialists

FAIRICUBE INTRODUCTION

Data science approach to solve environmental problems:

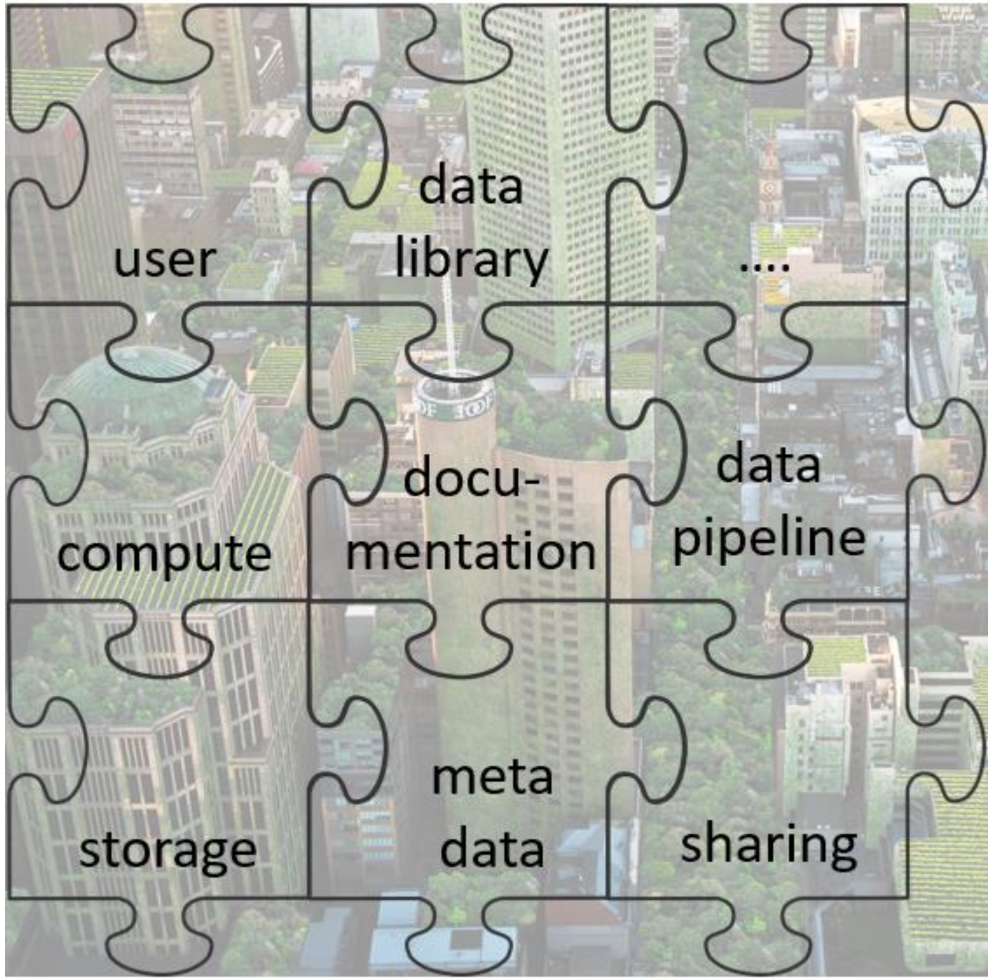
- All pieces need to be connected and interact with each other
- Additional puzzle pieces can be
  - Visualization
  - Programming language
  - Data bases
  - …



FAIRICUBE INTRODUCTION

Data science approach to solve environmental problems:

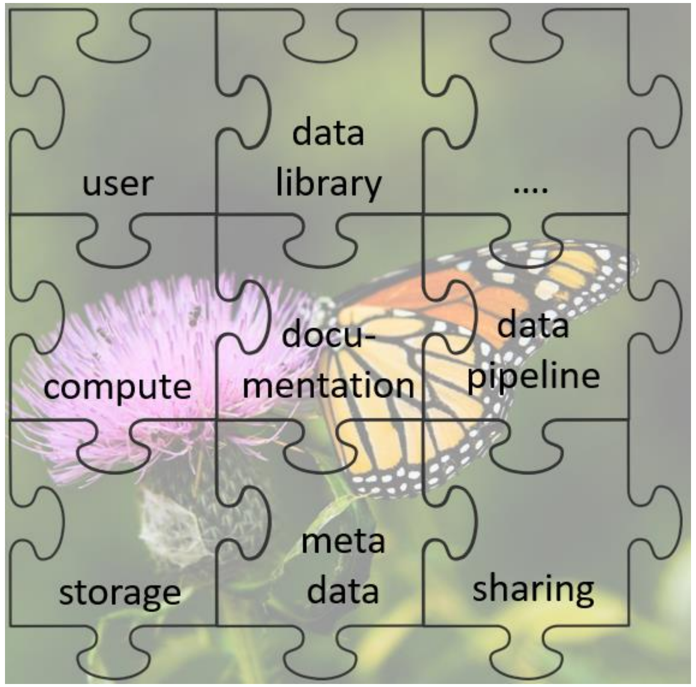
- Addressing EU green deal actions items focus on regular, gridded data
- Earth observation data
- Application of Machine Learning (ML) methods
- Example: urban adaptation to climate change (UC1) – EU GD Action Item Climate

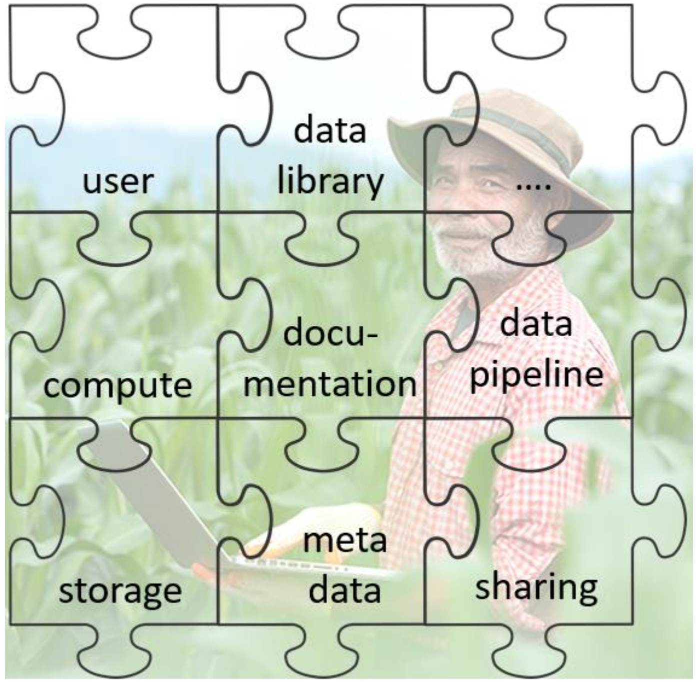


FAIRICUBE INTRODUCTION

EU GD Action Item being solved with large gridded EO data:

- EU GD Action Item Biodiversity (UC2, UC3 & UC5)
- EU GD Action Item Agriculture (UC2)
- EU GD Action Item Circular economy, Energy (UC4)

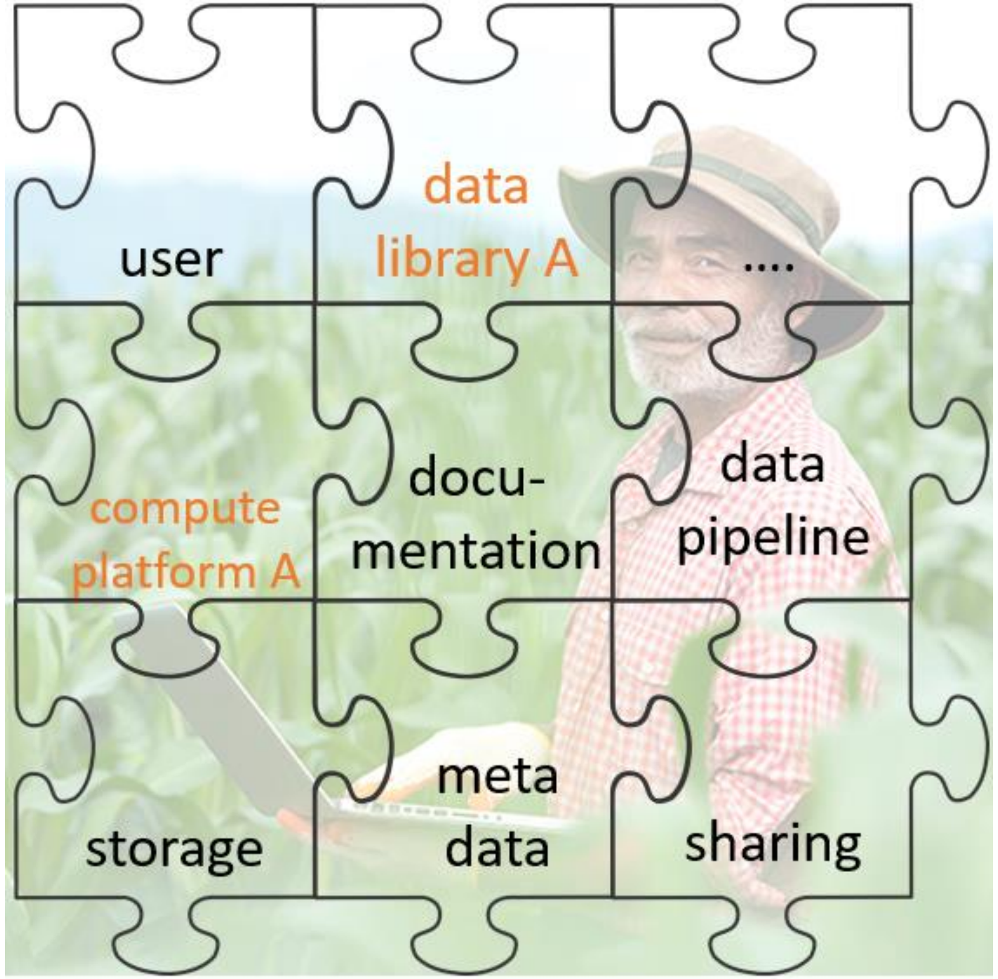




FAIRICUBE INTRODUCTION

Puzzle pieces:

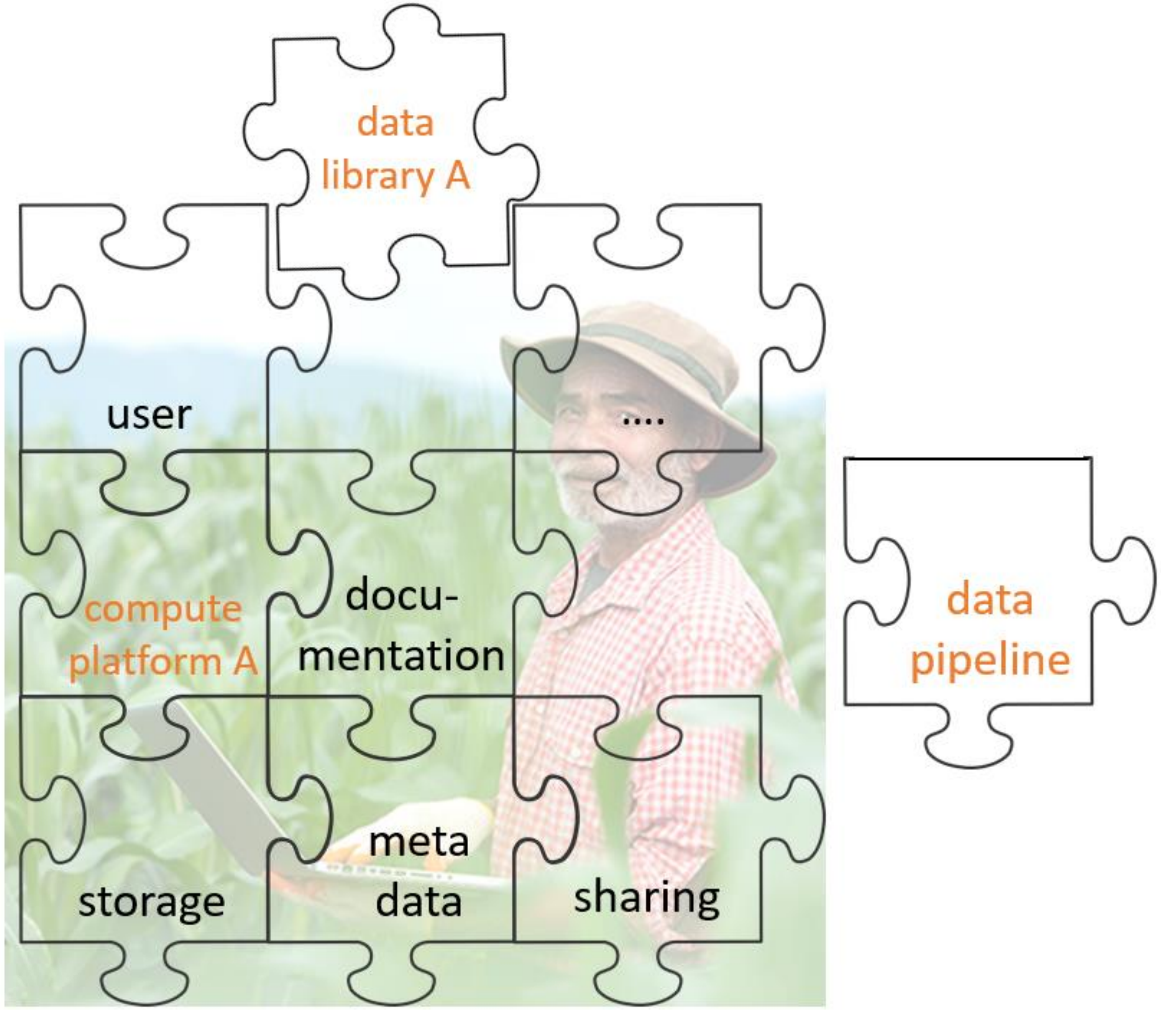
- Several if not all puzzle pieces can be exchangeable
- Local, remote or cloud resources



FAIRICUBE INTRODUCTION

Puzzle pieces:

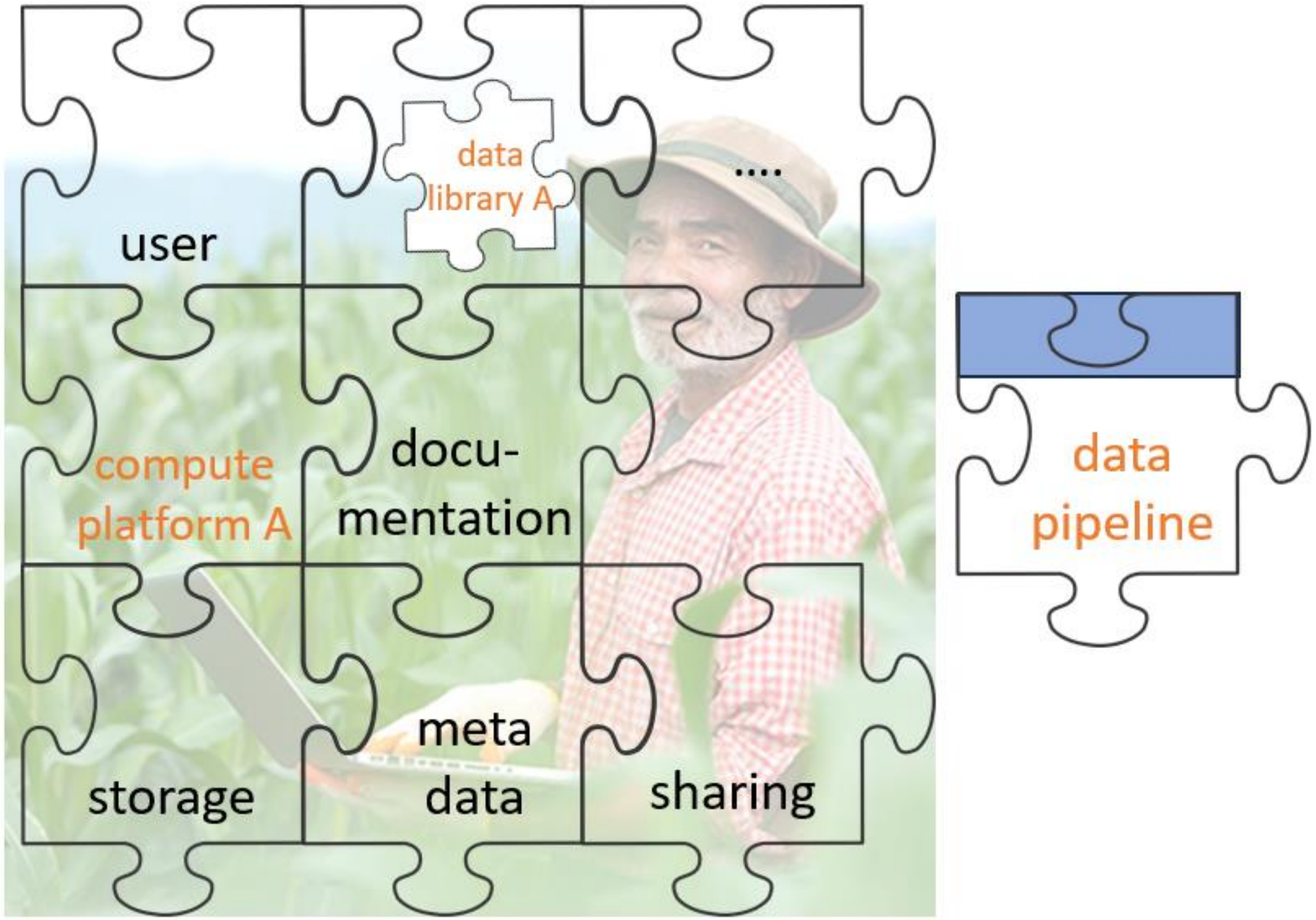
- Not all pieces fit into the puzzle
- Some have missing or different connectors
- Data might have different formats, CRS, gridding
- A lot of time is spent to work on this before any analysis work can be done



FAIRICUBE INTRODUCTION

Puzzle pieces:

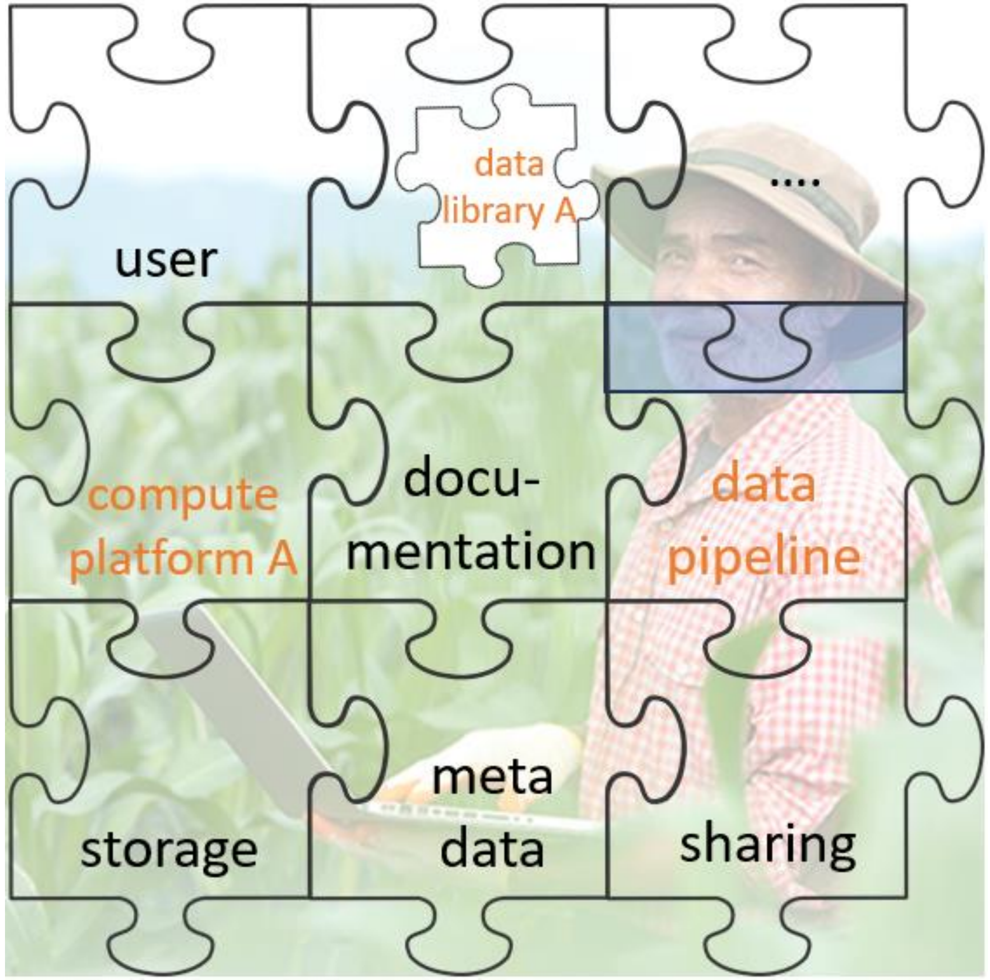
- FAIRiCUBE does not develop services but connectors (if missing)
- Complete metadata information
  - Data
  - Processing
- Documentation from start (data ingest) to finish (ML application)



FAIRICUBE INTRODUCTION

FAIRiCUBE Hub:

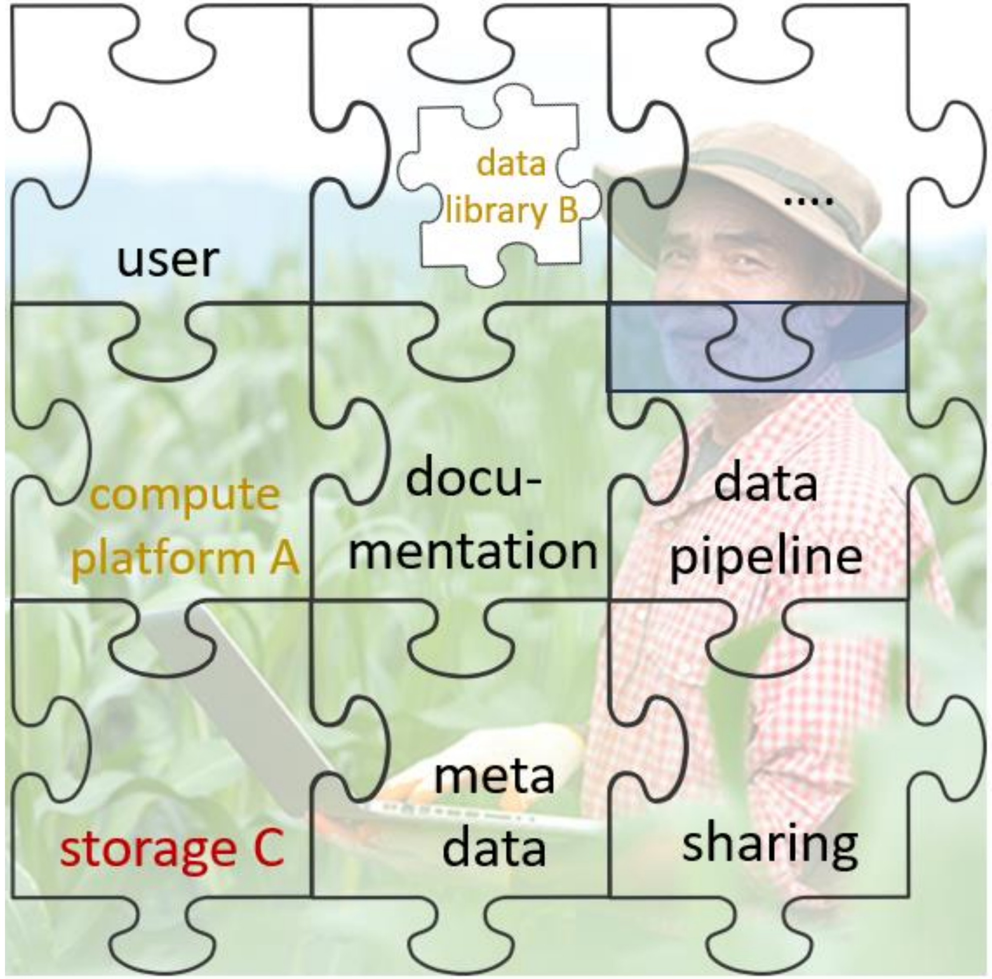
- Open and flexible workspace for data scientists
- F.A.I.R.
- Special focus on machine learning (ML) applications as data cubes are ML friendly



FAIRICUBE INTRODUCTION

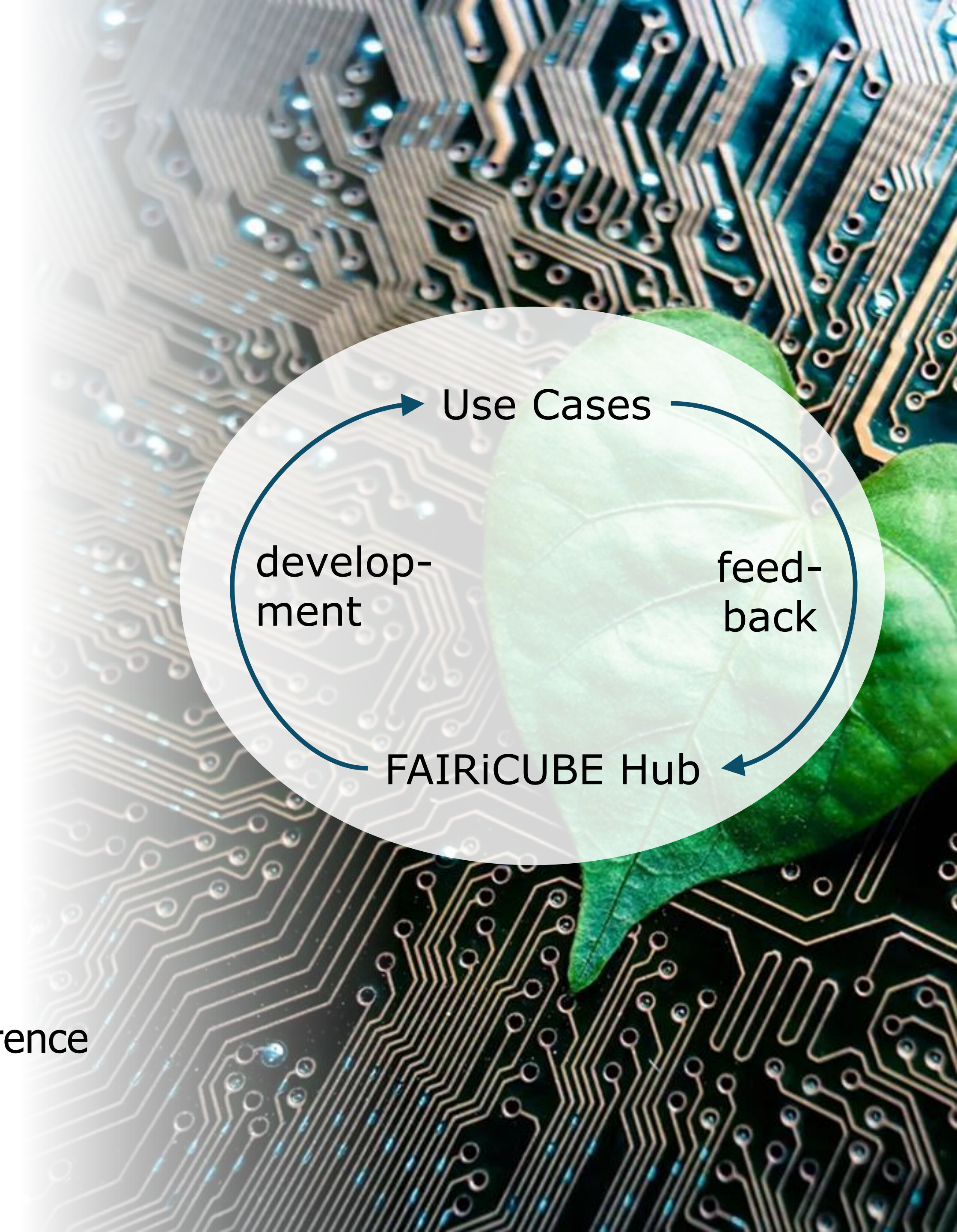
FAIRiCUBE Hub:

- Once connections are defined, infrastructure pieces can be exchanged to meet specific needs
- ML training: moderate amount of data but large compute needs
- ML inference: large amount of data but moderate compute needs
- Visualization: large amount of data but little compute needs



USE CASES

- 5 Use Cases being executed on the FAIRiCUBE Hub
  - urban / regional focus
  - Barcelona / Vienna / Oslo / Luxembourg
  - using different scale length, areas, objectives
  - research questions formulated
- UC1 Urban adaptation to climate change
- UC2 Biodiversity and agriculture nexus (regional)
- UC3 Analysis of genetic drift in Drosophila melanogaster (urban, biodiversity)
- UC4 Spatial and temporal assessment of neighbourhood building stock (urban, energy, circular economy)
- UC5 Validation of Phytosociological Methods through Occurrence Cubes (urban/regional focus, biodiversity)



USE CASE 2

THE BIODIVERSITY – AGRICULTURE NEXUS

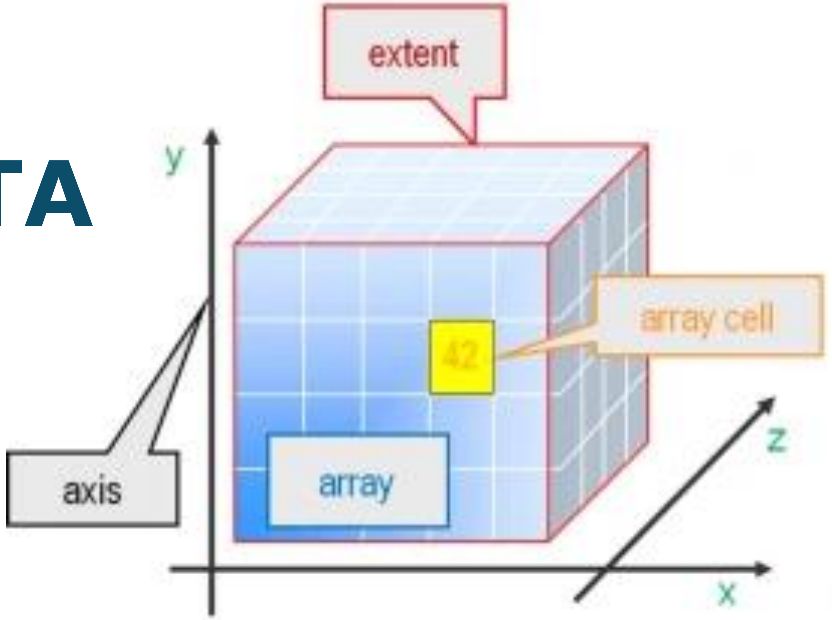
ABOUT THE USE CASE

- Studies the interaction between agriculture activities and biodiversity
- Aims at better-informed decision making in selecting more nature-inclusive practices promoting biodiversity
- UC objectives:
  - Investigate how farming activities affect biodiversity at the landscape level
  - Enhance understanding of how various agricultural practices impact the diversity of farmland bird species
  - Use Explainable AI and Causal ML to derive clear explanations of changes in the biodiversity
  - Store, analyze, and share data within FAIRiCUBE Hub enabling access for stakeholders in agriculture and biodiversity
  - Promote better informed decisions on and for the farm in a way that supports nature and wildlife

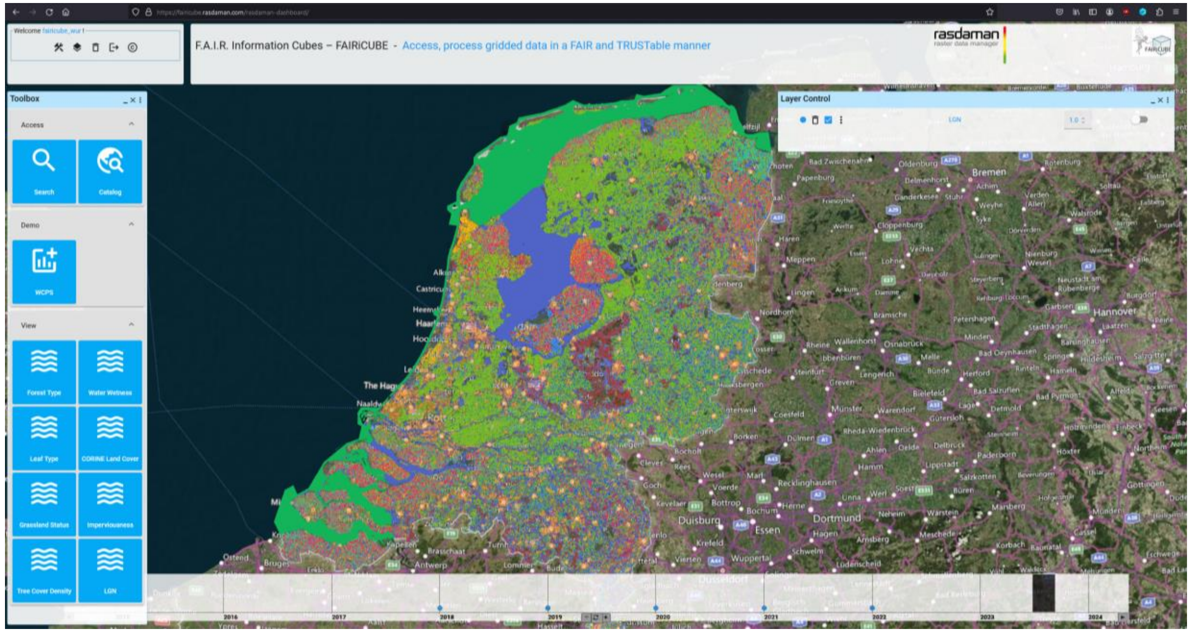
A small part of the Dutch agricultural landscape

EARTH OBSERVATION AND DATA CUBES INTRODUCTION

- Data cube as multi-dimensional array of values
- Geospatial Coverage DC – based on features / geospatial data with characteristics on spatial, temporal or spatiotemporal domain
- Earth observation DC – contain data from satellites, aircrafts, drones and ground sensors
- Platform for integrating earth observation data:
  - access the analysis ready EO data
  - create cubes from these data
  - performs query and analysis
  - integrate own data and algorithms
  - customize data and processing pipeline
  - connect own tools
  - use a pre-prepared template of a use case
  - collaborate with other user



https://doi.org/10.1186/s40537-020-00399-2

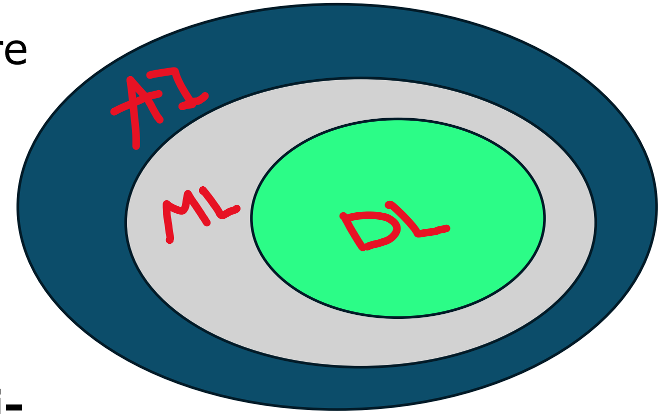


https://fairicube.rasdaman.com/

ARTIFICIAL INTELLIGENCE STARTER

AI vs ML vs DL

- Artificial Intelligence **(AI)** is the broad concept of creating machines that can perform tasks which require human intelligence.
- Machine learning, deep learning, symbolic reasoning, expert systems, etc., are some techniques being used as part of AI goal.



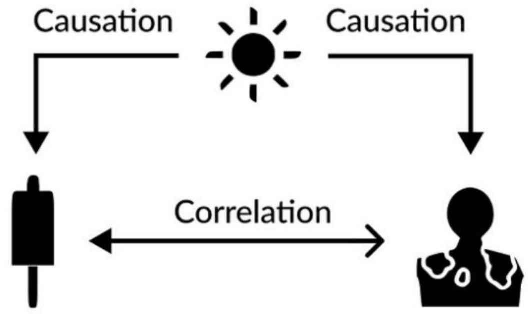
- Machine Learning **(ML)** focuses on the development of algorithms and statistical models that learn patterns and make predictions or decisions from experience/observation without being explicitly programmed.
- ML techniques include **supervised learning**, **unsupervised learning**, **semi-supervised learning**, **reinforcement learning**, **self-supervised** and more.

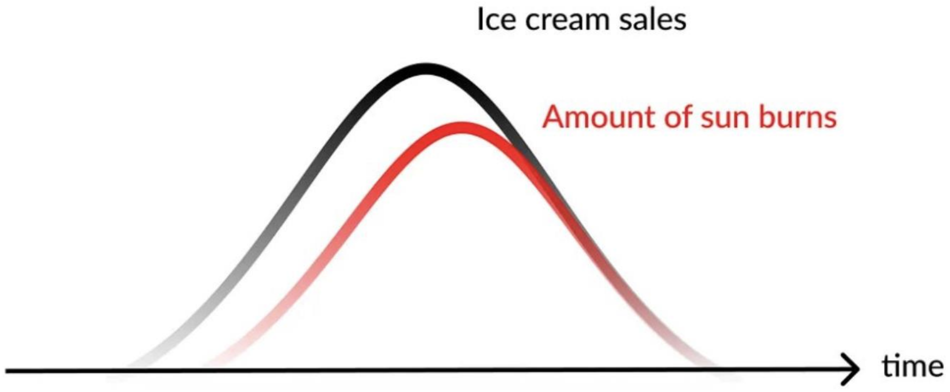
- Deep Learning **(DL)** is a subset of ML that uses artificial neural networks with multiple layers (the term "deep") to model and learn complex patterns in large datasets.
- DL algorithms automatically learn hierarchical representations of data, which extracts features from the input data.

CORRELATION IS NOT CAUSATION

- Machine learning is a very powerful **correlation-pattern recognition** system that works well on data from the same distribution that it was trained on — but it is incredibly vulnerable to distribution changes.

- Causal machine learning aims to go beyond correlation and determine whether **one variable directly influences another** and, if so, how much.

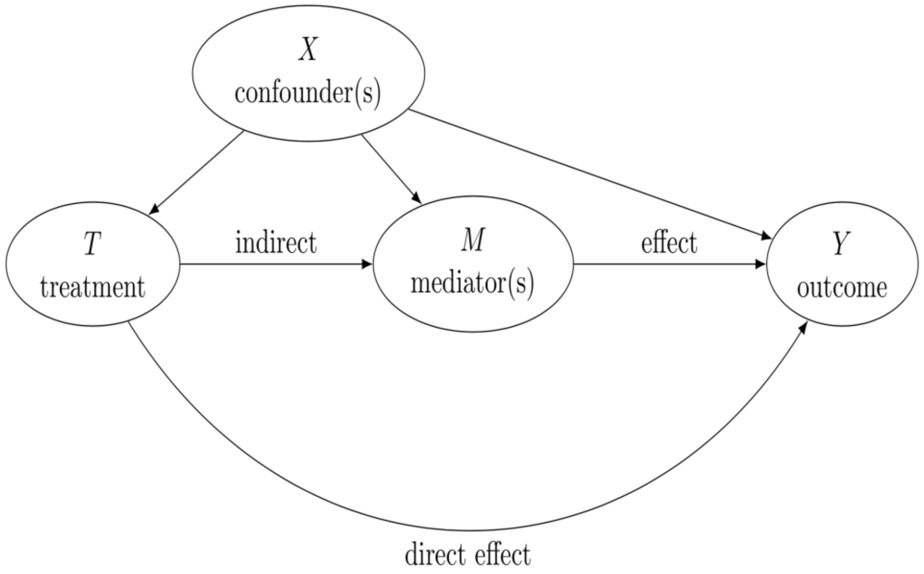




CAUSAL AI METHODS

Causal Discovery: Identifies causal relationships or causal structures from observational data

- Graphical models
- Constraint-based methods
- Score-based methods



Causal Inference (causal effect estimation):
assuming a causal relationships is already in a causal model and make predictions about the effect of interventions or treatments

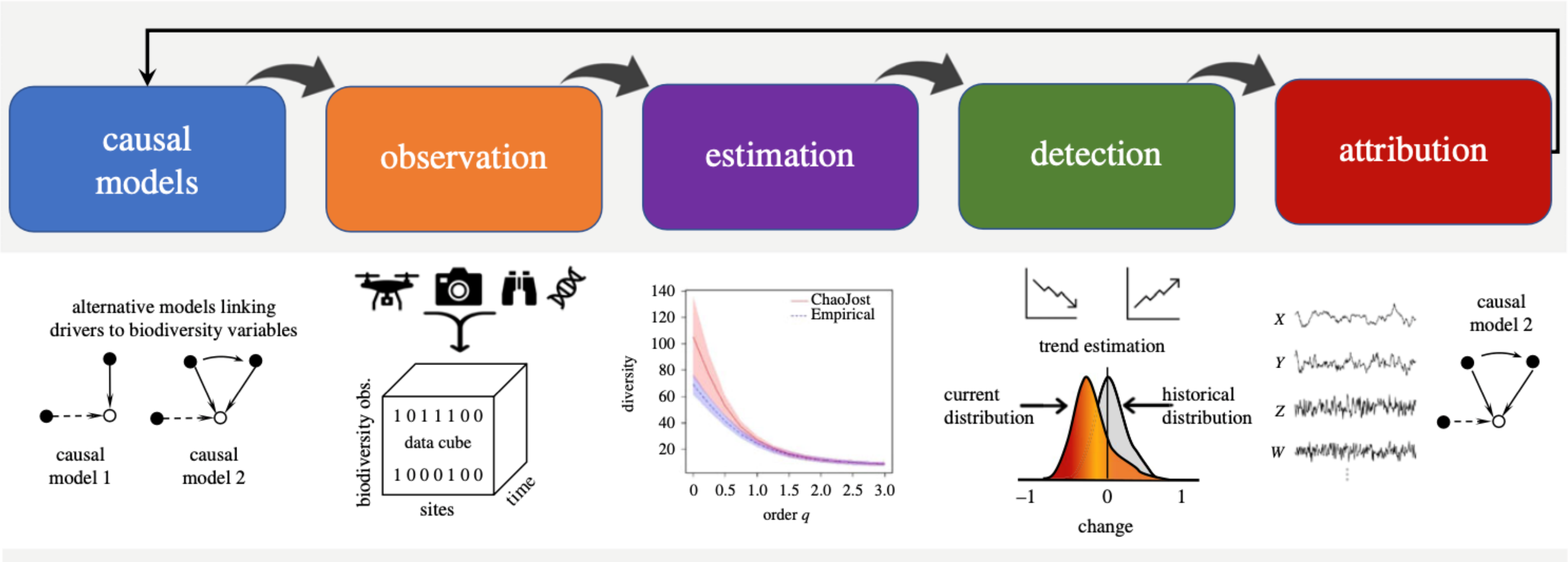
Counterfactual Analysis (what-if scenarios): a hypothetical scenario where one or more factors are changed while keeping everything else the same. The aim is to determine how an output will change in a counterfactual scenario.

CAUSAL AI CHALLENGES

- **Confounding variables:** Factors that affect both the treatment and the outcome, may lead to false correlations
- **Selection bias:** When the sample is not representative of the population, leading to biased estimates
- **Establishing causality:** Distinguishing between correlation and causation requires careful study, analysis and domain knowledge.

USE CASE FRAMEWORK

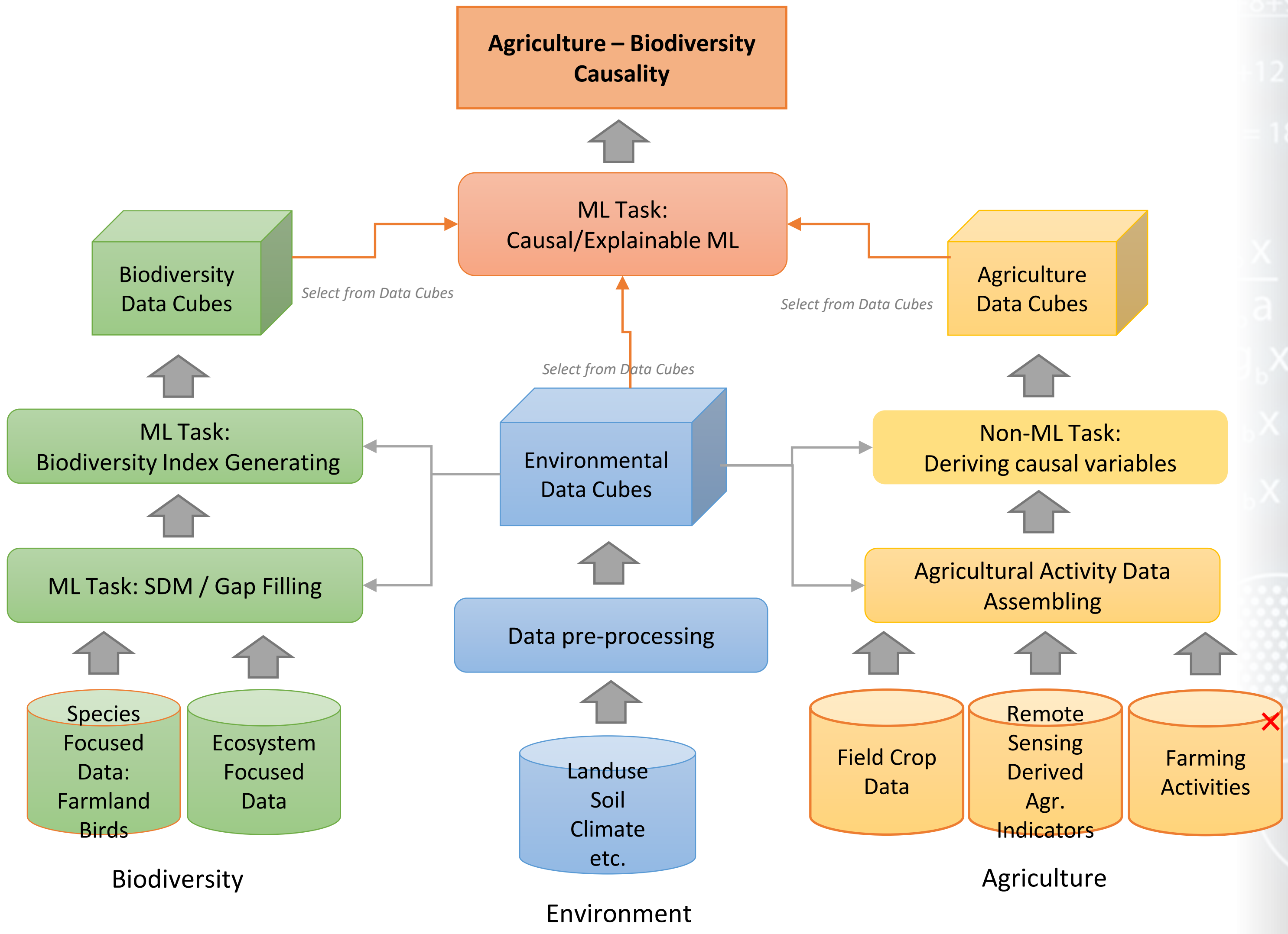
Gonzales et al. (2023)[1]: “A formal framework and guidelines for the detection and attribution of biodiversity change is needed.”



“Biological diversity is a measure of life’s compositional variation across different levels of organization—genome, population, species, and ecosystems—and its changing state over dimensions of space and time.“

[1]Gonzalez, A., Chase, J., & O’Connor, M. (2023). A framework for the detection and attribution of biodiversity change. *Philos Trans R Soc Lond B Biol Sci*, *378*(1881), 20220182. https://doi.org/10.1098/rstb.2022.0182

USE CASE WORKFLOW



QUESTIONS?

FEEDBACK SESSION
MODEL RESULTS, VARIABLES AND DATA

DATA COLLECTED AND PROCESSED SO FAR (FOR FILLING DATA CUBES)

- Agriculture – obtained from AgroDataCube
- Environmental – open data
- Biodiversity – species observations from NDFF – farmland birds
  - pre-processed: filtering, abundance calculation and gridding

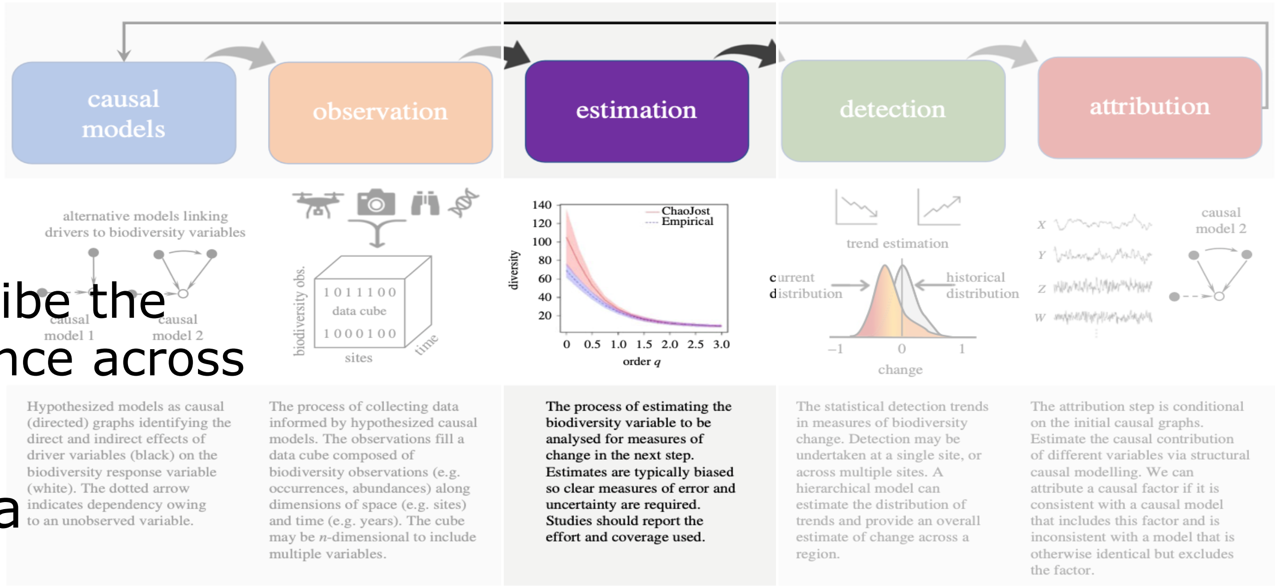
Agriculture	Environmental	Biodiversity
Crop rotation	Land use	Farmland birds abundance
No. of grassland mowing events	NDVI (EO based)	
Annual greenness	Objects height (DEM based)	
Winter greenness	Soil type	
Woody landscape elements		

QUESTIONS

1. Which variables (biodiversity, agriculture, environment) are missing?
2. What data is available for these variables, and who can provide it?

SPECIES DISTRIBUTION MODELLING (SDM)

- MaxEnt software is based on the maximum-entropy approach for modelling species distributions.
- MaxEnt works with three primary inputs:
  - The location of known presence points
  - A study area
  - Explanatory variables, or covariates, that describe the environmental factors that may relate to presence across the study area
- The study area (background points) defines a landscape where presence is possible.



SDM DATA

- At the time of modelling with MaxEnt, we only had access to, and hence used, the following data

Covariates → Agriculture, Environmental; Presence-only observations → Biodiversity

Agriculture	Environmental	Biodiversity
Crop rotation	Land use	Farmland birds abundance
No. of grassland mowing events	NDVI (EO based)	
Annual greenness	Objects height (DEM based)	
Winter greenness	Soil type	
Woody landscape elements		

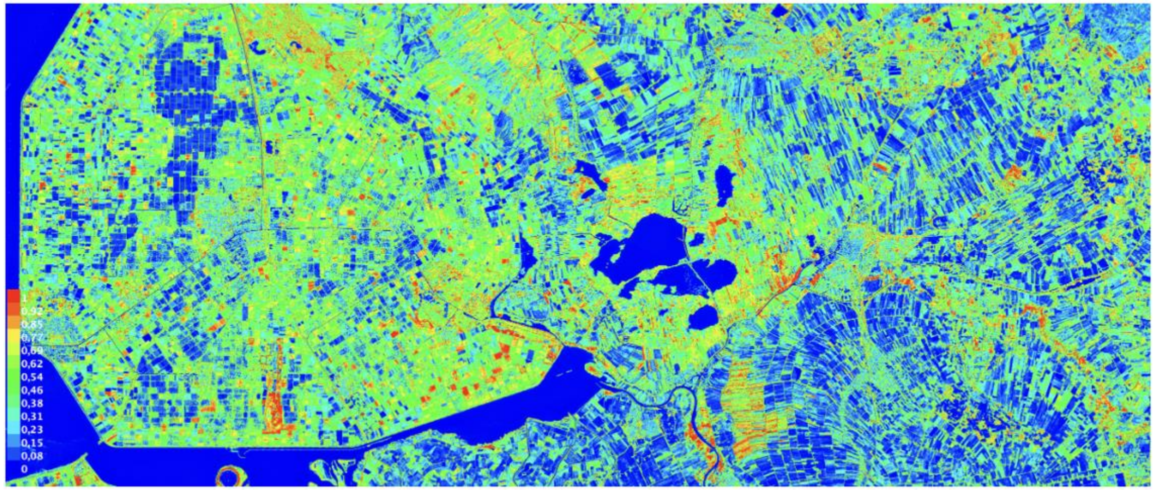
SDM RESULTS

- The table shows a general overview of bird species analysis generated by MaxEnt. An average of ten replications is used to calculate the results.
- For some points we had very few samples, so we cannot trust the prediction. Usually accuracy is quite high in these cases due to *overfitting*.

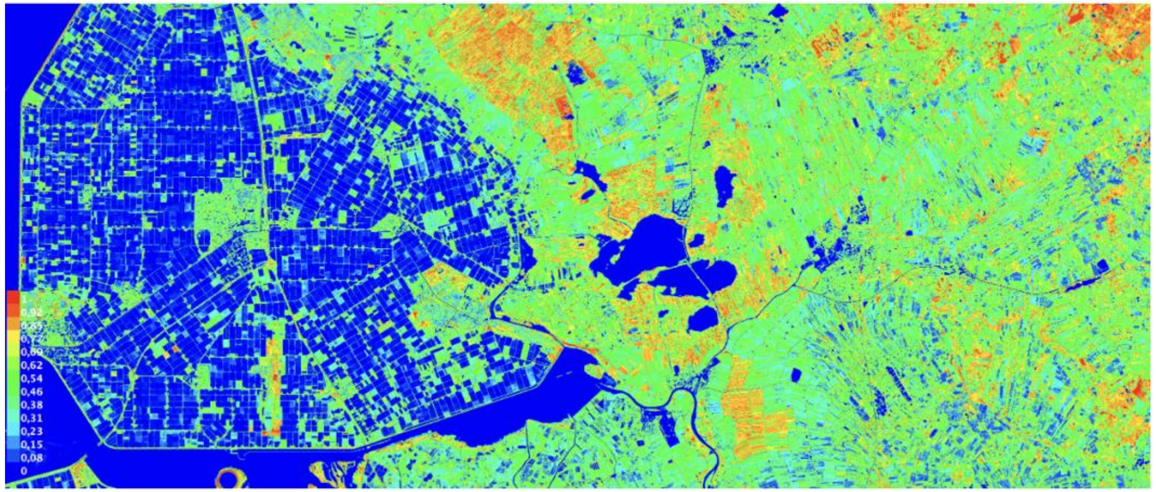
Species	AUC	stddev	#samples for train	#samples for test	#points used to determine distribution
Veldleeuwerik	0.74	0.03	883	98	10883
Graspieper	0.71	0.01	1347	150	11342
Putter	0.79	0.03	418	47	10418
Kwartel	0.81	0.05	124	14	10124
Geelgors	0.78	0.03	573	64	10573
Watersnip	0.93	0.02	98	11	10098
Scholekster	0.78	0.02	745	83	10744
Grutto	0.84	0.01	2260	252	12259
Gele kwikstaart	0.79	0.01	1528	170	11523
Wulp	0.81	0.06	106	12	10106
Ringmus	0.83	0.02	357	40	10356
Spreeuw	0.71	0.02	595	67	10593
Grasmus	0.76	0.02	862	96	10861
Tureluur	0.85	0.06	148	16	10148
Kievit	0.74	0.01	1611	179	11609

SDM RESULTS

- Maps generated by MaxEnt for two birds: Putter(left picture) and Geelgors (right picture).
- Warmer colour shows a higher probability of observing the specie.



Putter



Geelgors

SDM CONCLUSIONS

- The **advantage** of using MaxEnt:
  - Working with presence-only data
  - It includes interaction between covariates (environmental variables)
  - We wanted to have some base model at the beginning.
  - The predictions can help us with gap filling.

- However, there are some **limitations:**
  - For some species we do not have enough samples, even with regularization the model can overfit.
  - In addition, the precise locations (longitude and latitude) of species (the csv file) are not often accessible.
  - Furthermore, scaling MaxEnt models is difficult, it is time consuming to run the model for more than one species.

BIODIVERSITY – AGRICULTURE CAUSALITY

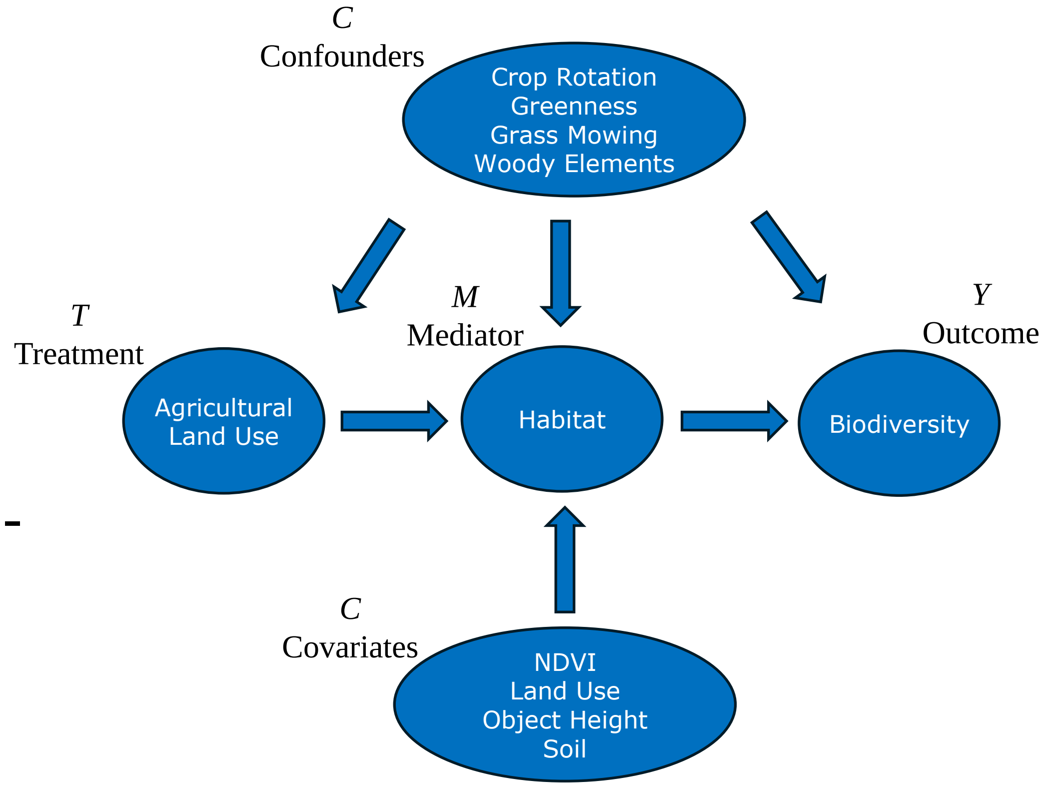
- Use of **Causal ML** to attribute biodiversity changes to specific agricultural practices
  - Not only predict the correlations but also provide meaningful explanations for those predictions
  - Enhancing the overall interpretability of the modelling results
- **Counterfactual Analysis** – what would be an impact on biodiversity if we change agriculture management practice (what-if scenario)

BIODIVERSITY – AGRICULTURE CAUSALITY

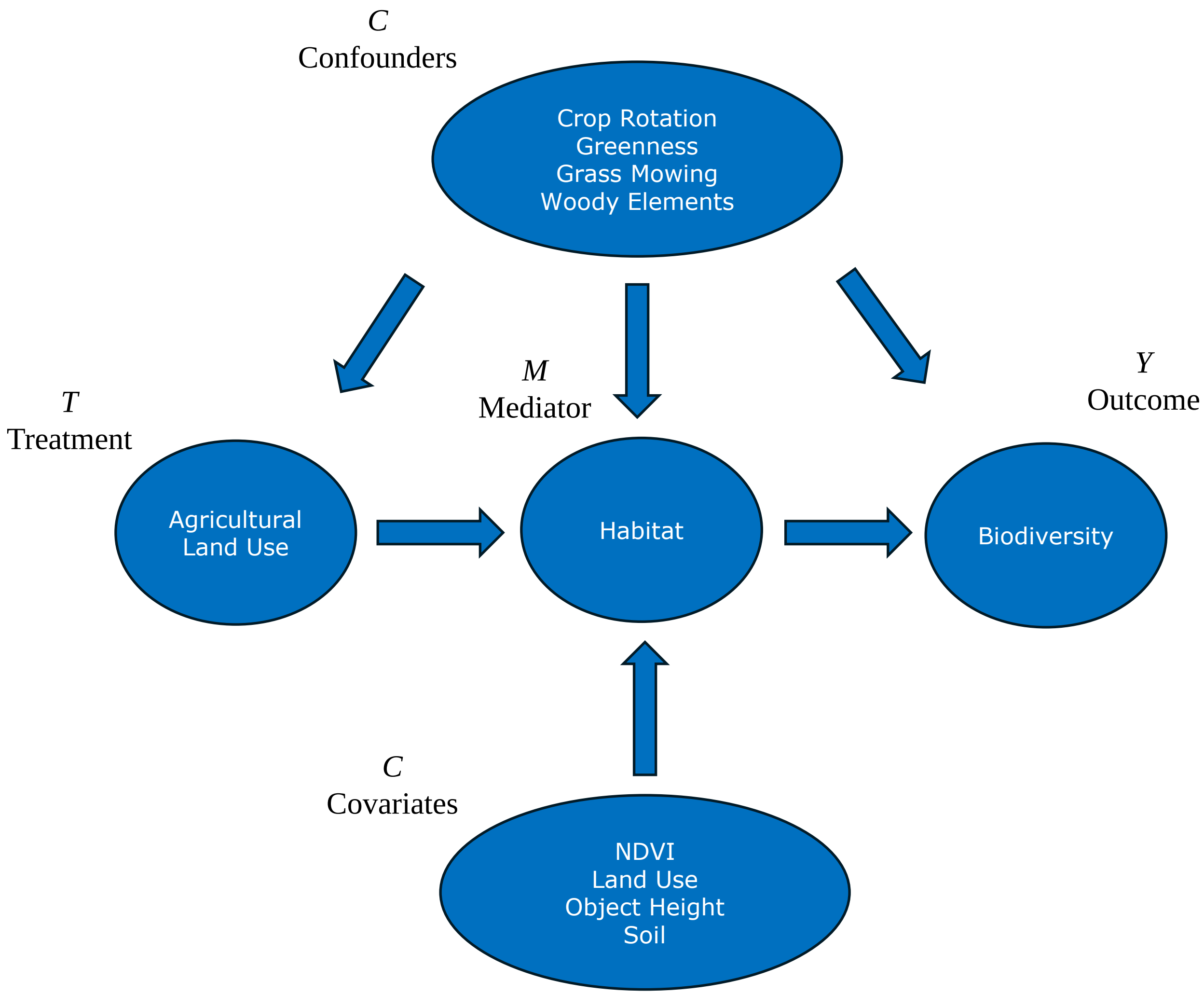
FAIRiCUBE

Causal diagram – directed graph, displays causal relationships between variables in a causal model

- **Agricultural land use** (Treatment) – arable land or grassland makes unique habitat for farmland birds
- **Habitat** (Mediator) – environment occupied by bird species
- **Agricultural management practices** (Confounders) - variables which differ in different scenarios and influences land use, habitat and biodiversity
- **Environmental variables** (Covariates) – stable elements - does not change in different scenarios
- **Biodiversity** (Outcome) – Estimated measures of biodiversity based on species abundance



BIODIVERSITY – AGRICULTURE CAUSALITY



QUESTIONS

- Is the relationship between agriculture and biodiversity suitably described by the causal graph?
- What other data or graph elements in the Use Case are missing?

QUESTIONS?

FEEDBACK SESSION

WHAT ARE POSSIBLE APPLICATIONS?

Join at menti.com | use code 8918 4770

ZONING

- The results of the Observation and Estimation steps for biodiversity can be used to categorize agricultural landscapes and e.g. administrative regions, based on predicted suitability.

- For example:
  The agricultural landscape in Gelderland has a 30% higher predicted suitable for the 'Geelgors than the Flevopolder.

WHAT-IF?

- Causal modelling allows reasoning about counterfactual situations to answer 'What-if?" type of questions.
- For example:
  What is the predicted change
  in biodiversity for farmland bird species
  XXXX when the woody landscape elements i
  n agricultural region YYYY
  are increased by 10%?

SMART TOOLS

- The presented approach aims at improved understanding of causalities between farm activities and changes in biodiversity.
- When results are sufficiently robust, the model could be incorporated into advisory tools for farmers or policy makers, to help assess the consequences of actions.

OTHER FEEDBACK

- FEASIBILITY
- ETHICS
- ROBUSTNESS
- FAIRICUBE HUB

Join at menti.com | use code 8918 4770

FEASIBILITY

- On a scale of 1 (fully disagree) to 5 (agree completely) how do you rate the feasibility of the presented use case approach on the following aspects?
  - Sufficient data is available to make it work
  - Causal modelling is a good approach for nature studies
  - The model can be expected to produce usable outcomes

ETHICS

- The presented approach has many opportunities for introducing biases, e.g. from observational data or from missing important confounding variables in the causal network. It also tries to deal with sensitive topics, i.e. biodiversity decline and agricultural practices.

  Considering such aspects how would you rate the ethical risks of potential applications on a scale of 1 (not relevant at all) to 5 (very relevant)?

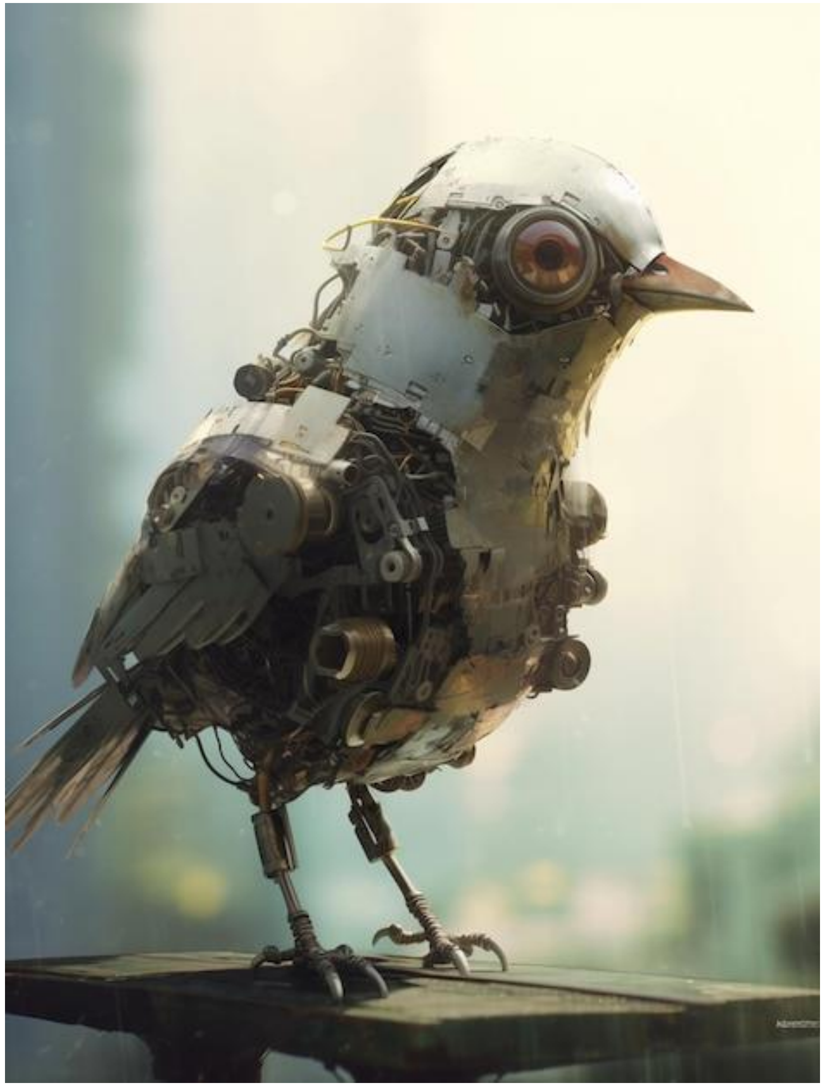
ROBUSTNESS

- Please indicate (between 1 – 100%) how statistically robust the outputs of the causal model should be in order to be useful to you.

  For example: if the casuality between changes in farmland suitability for bird species XXXX and annual farmland greenness can be proven, the robustness of this should at least be 60% to be valuable (to you).

FAIRICUBE HUB

- When the FAIRiCUBE Hub has been sufficiently developed, which of the following statements would apply to you (you can select multiple)?



CAN ARTIFICIAL INTELLIGENCE SUPPORT BIRD CONSERVATION?

WAGENINGEN
UNIVERSITY & RESEARCH

FAIRiCUBE

OPEN DISCUSSION

THANK YOU FOR PARTICIPATING!

- Second seminar
- Feedback on this seminar
- Can we publish the recording for dissemination purposes?
- Our newsletter

Funded by
the European Union

This project has received funding from the Horizon Europe research and innovation programme under grant agreement No. 101059238.