

The KI:STE platform

Easing and promoting ML and AI usage in earth system science

Objectives

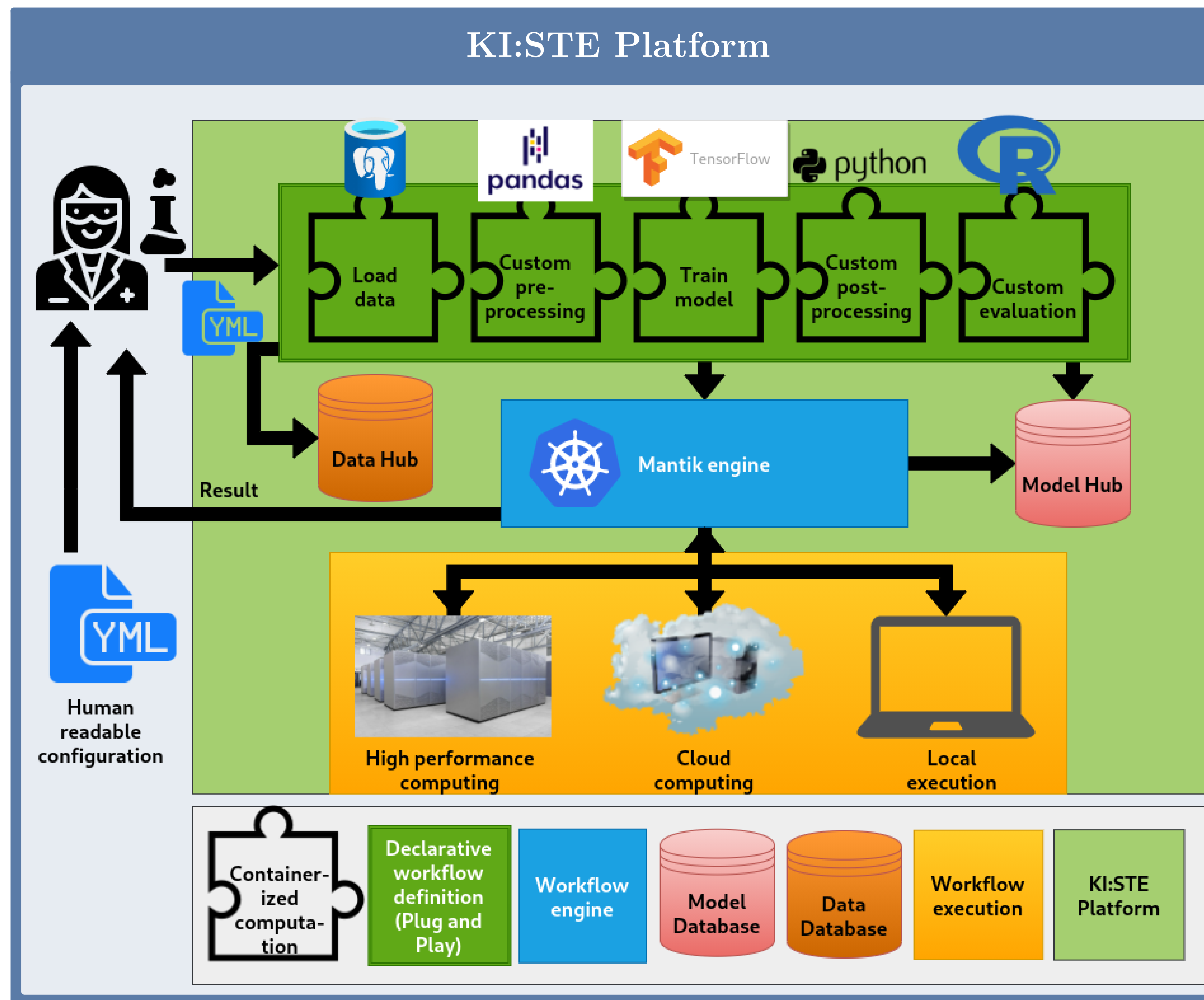
- **Easy usage** with a handy front-end.
- **Sharing** is caring by Open Source Software and model hub.
- **Extensibility** to add your own libraries.
- **Reproducibility** at full strength.
- **Comparability** of models by a recommender.
- **Compute everywhere**: Execution on various infrastructures.

Features

- **Cross-framework support**: Containerized for support of the major AI tools.
- **Plug and Play**: Choose from pre-configured execution steps.
- **Collaboration**: Share versions of models connected to data.
- **Language agnostic**: Write your own applications in any language you like.
- **Declarative workflow definition**: Human readable and easy to share.
- **Data projects**: Use curated and version controlled data.

Human Readable Configuration

```
name: multiply
bridge: mantik/sklearn.simple
kind: algorithm
type:
  ...
```



Data Projects

- Generic adapters for **common data formats**
- Adapters for most-used datasources (e.g. Copernicus)
- **Versioning** and **snapshotting** of data for full reproducibility
- Execution environment aware data provisioning

Execution Environments

- **HPC**: Developed for JUWELS@JSC Juelich.
- **AWS**: Currently deployed on AWS.
- **Other**: Helm Charts for different clouds.
- **Local execution**: Mini-Mantik, running in minikube.

Supported Frameworks

We currently support: tensorflow, sklearn, pandas, geopandas, SQL adapter, TOAR data loader

Coming Soon

- **Open source** publication of the platform software.
- **Your frameworks** will be integrated.
- Meta learning and **recommendation system**.
- Five PhD **research projects** using the KI:STE platform.
- **Data versioning** and data distribution.
- **Model gallery**.

Links

Find further information on kiste-project.de.

Contributors

- [Thomas Seidler^c](#)
- [Christian Autermann^a](#)
- [Dr. Benedikt Gräler^a](#)
- [Dr. Scarlet Stadler^b](#)
- [Dr. Markus Abel^{cd}](#)
- [Dr. Markus Quade^{cd}](#)
- [Norbert Schultz^{cd}](#)
^a52°North GmbH
^bJSC Juelich
^cAmbrosys GmbH
^dMantik UG

Funded by the German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety, grant number **67KI2043F – KISTE**.

The KI:STE platform

Easing and promoting ML and AI usage in earth system science



[Abstract](#)

[The KISTE project](#)

[The KISTE platform](#)

[Objectives](#)

[Features](#)

[Partners](#)

[Home](#)

[Section](#)

[Next](#)

[Previous](#)

[Contents](#)

The KI:STE platform

Easing and promoting ML and AI usage in earth system science

Abstract

Earth system modeling is virtually impossible without dedicated data analysis. Typically, data are big and due to the complexity of the system, adequate tools for the analysis lie in the domain of machine learning or artificial intelligence. However, earth system specialists have other expertise than developing and deploying state-of-the-art programming code which is needed to efficiently use modern software frameworks and computing resources. In addition, Cloud and HPC infrastructure are frequently needed to run analyses with data beyond Tera- or even Petascale volume, and corresponding requirements on available RAM, GPU and CPU sizes.

Inside the KI:STE project (www.kiste-project.de), we extend the concepts of an existing project, the Mantik-platform (www.mantik.ai), such that handling of data and algorithms is facilitated for earth system analyses while abstracting technical challenges such as scheduling and monitoring of training jobs and platform specific configurations away from the user.

The principles for design are collaboration and reproducibility of algorithms from the first data load to the deployment of a model to a cluster infrastructure. In addition to the executive part where code is developed and deployed, the KI:STE project develops a learning platform where dedicated topics in relation to earth system science are systematically and pedagogically presented.

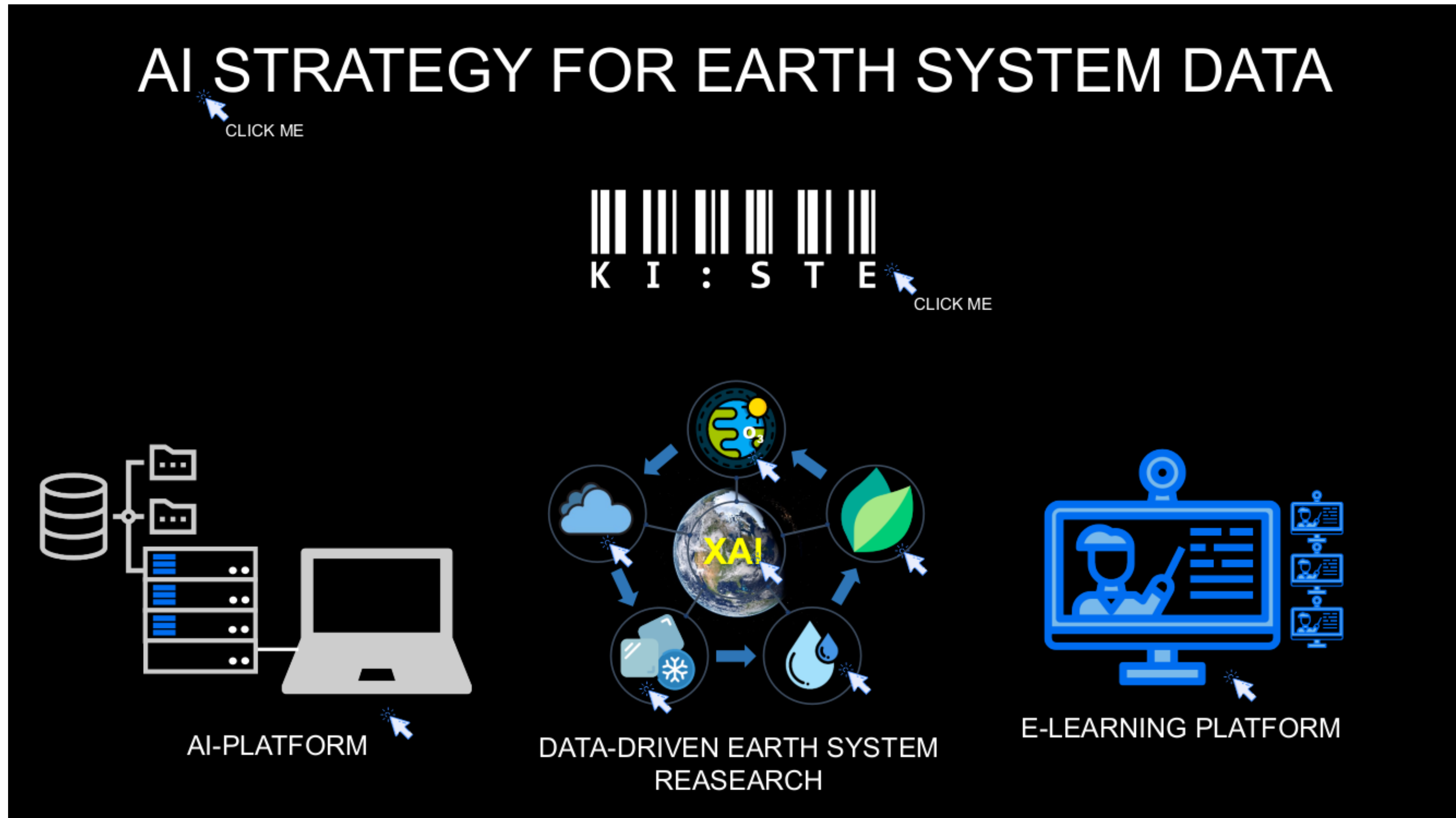
In this presentation, we show the architecture and interfaces of the KI:STE platform together with a simple example.

The KI:STE platform

Easing and promoting ML and AI usage in earth system science

The KISTE project

Visit Scarlet Stadtler's talk!



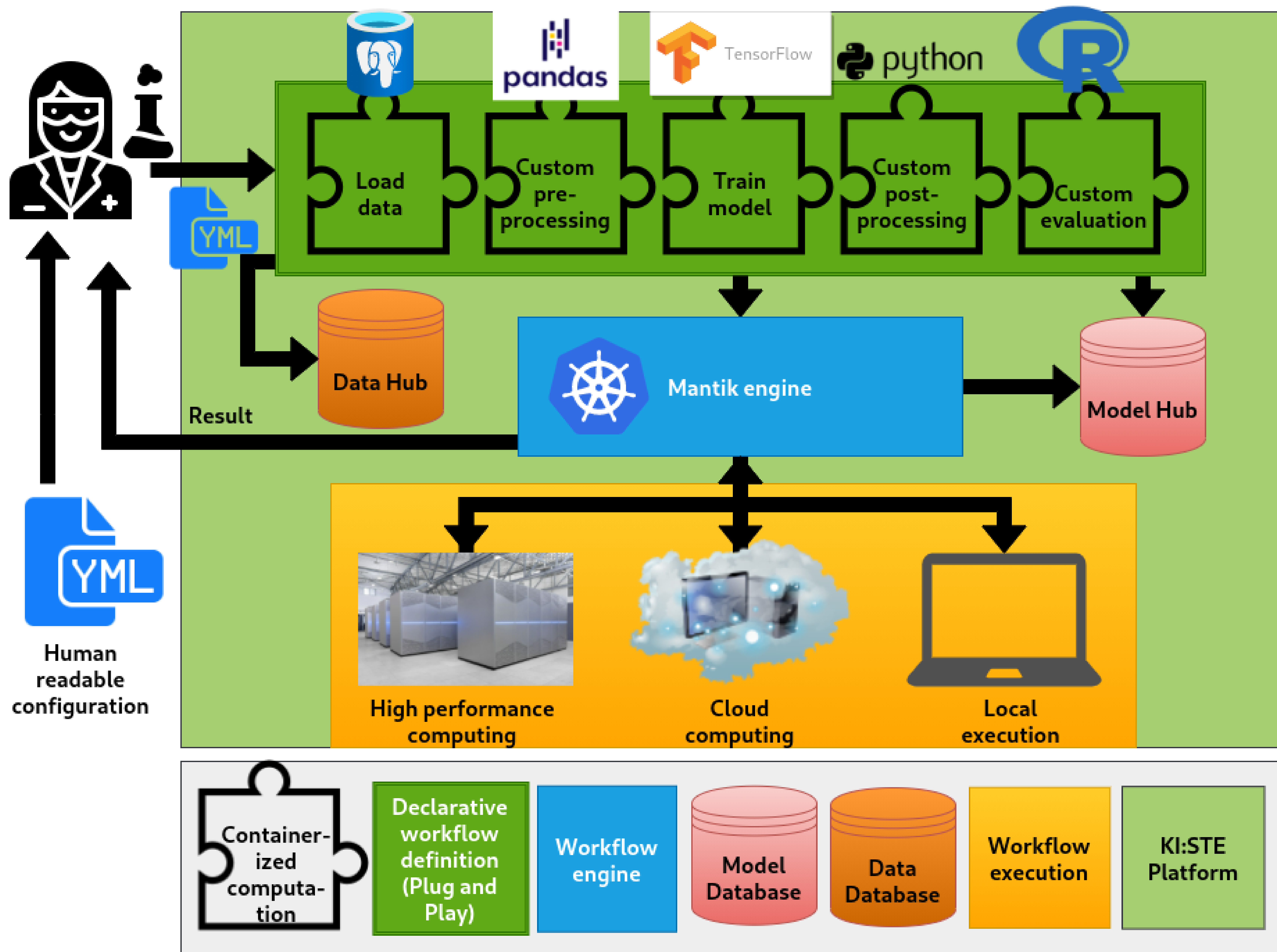
The slide features a black background with white and blue text and graphics. At the top, the title "AI STRATEGY FOR EARTH SYSTEM DATA" is displayed in large white letters, with a blue arrow pointing to the word "AI" and the text "CLICK ME" below it. Below the title is a white barcode with the letters "K I : S T E" underneath. A blue arrow points to the letter "E" with the text "CLICK ME" below it. The central graphic consists of a globe with "XAI" written on it in yellow. Six blue arrows radiate from the globe to icons representing different earth system components: a cloud, a leaf, a water drop, a snowflake, a server rack, and a person. Below this central graphic are three main sections: "AI-PLATFORM" on the left, which includes icons of a database, server racks, and a laptop; "DATA-DRIVEN EARTH SYSTEM RESEARCH" in the center, which is the central graphic; and "E-LEARNING PLATFORM" on the right, which includes an icon of a computer monitor displaying a person and a list of items.

EGU21-211 | vPICO presentations | ITS4.4/AS4.1/CL5.2.8/ESSI1.1/NP4.4

The KI:STE platform

Easing and promoting ML and AI usage in earth system science

Sketch of the KISTE platform

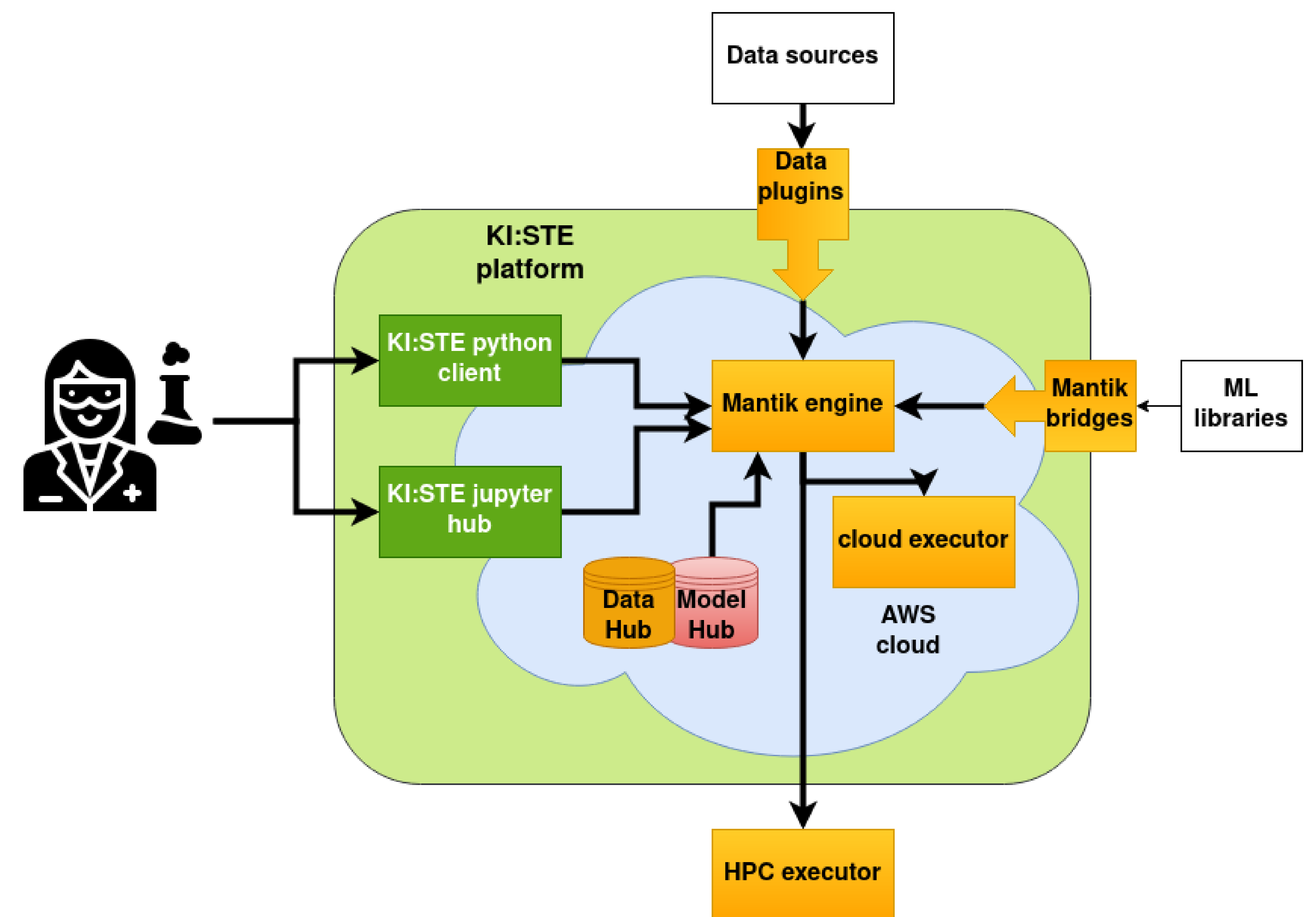


The KI:STE platform

Easing and promoting ML and AI usage in earth system science

Platform Architecture

- The user can access KISTE platform via our jupyter hub or the python client / CLI.
- The backend (Mantik engine) is currently deployed on AWS.
- A workflow definition is translated into a computational graph using our Mantik and data plugins ("bridges"). All bridges are containerized and can be configured.
- The computational graph is executed in the cloud with our Docker executor or on HPC with our Singularity executor.



The KI:STE platform

Easing and promoting ML and AI usage in earth system science

Objectives

- **Easy usage** with a handy front-end. Currently we use a jupyter hub as frontend. We will implement our own web-based frontend, where models, data and available bridges can be viewed in one place.
- **Sharing** is caring by Open Source Software and model hub. Mantik will be released on July 1st with the KISTE platform setup coming along.
- **Extensibility** to add your own libraries. We are working hard to provide access to the most common ML libraries. If you want to add something you can easily build your own bridge.
- **Reproducibility** at full strength. We provide curated and version controlled datasets. Models are stored and version controlled in our model hub, based on the version control of mantik bridges.
- **Comparability** of models by a recommender. Meta-learning on most common model configurations is implemented to recommend the best model class to you based on available features and target data.
- **Compute everywhere**: Execution on various infrastructures. Mantik executes natively wherever docker or singularity containers can be run.

The KI:STE platform

Easing and promoting ML and AI usage in earth system science



Features

Cross-framework support

Plug and Play

Reproducibility

Language Agnostic

Declarative workflow definition

Data Projects

[Home](#)

[Section](#)

[Next](#)

[Previous](#)

[Contents](#)

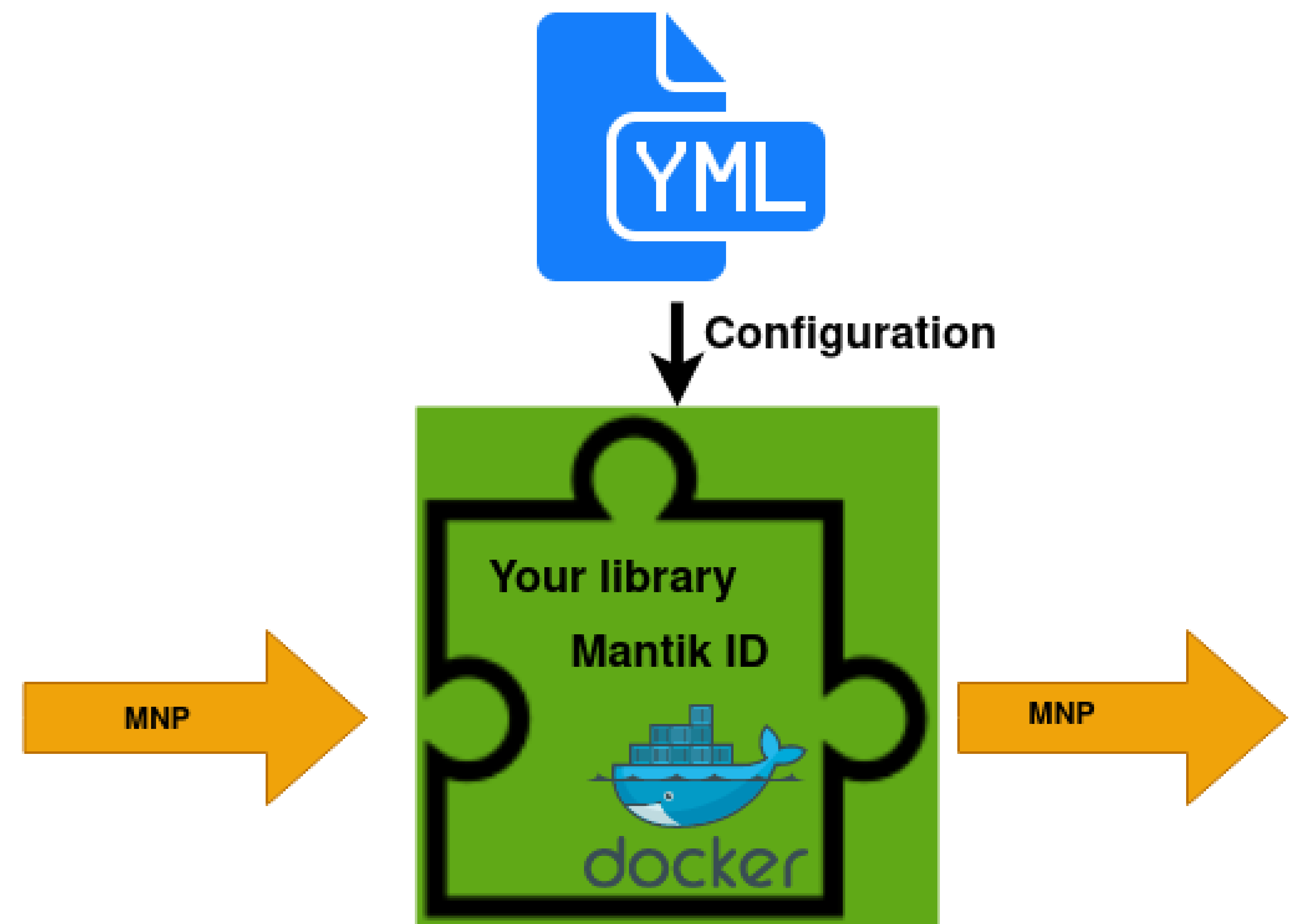
The KI:STE platform

Easing and promoting ML and AI usage in earth system science

Cross-framework support and language agnosticism

- The KI:STE platform uses **Mantik bridges**: Libraries are encapsulated in Docker or Singularity containers.
- Bridges have input and output type definitions.
- Bridges communicate via **Mantik Node Protocol (MNP)**, a gRPC¹ based protocol specification.
- MNP clients (stubs) are generated automatically. The Mantik engine can handle every programming language supported by gRPC.

¹gRPC is a Remote Procedure Call framework that mediates communication between microservices.

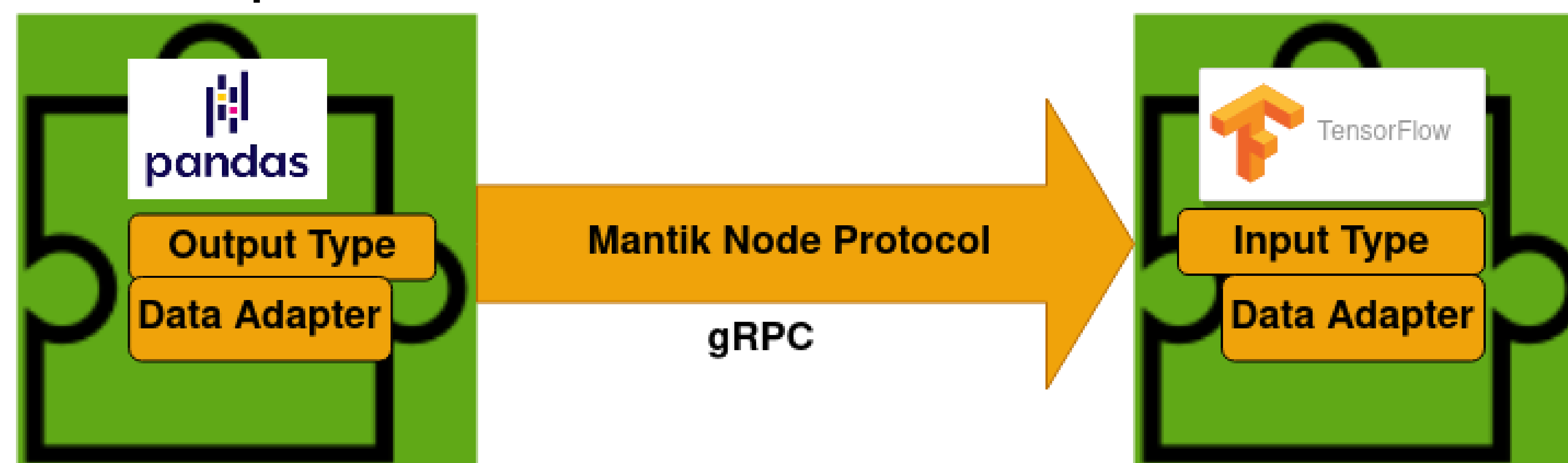


The KI:STE platform

Easing and promoting ML and AI usage in earth system science

Plug and Play

- Mantik bridges can be freely combined to a workflow.
- Due to the typisation of the bridges, Mantik engine knows how to handle in- and output data.
- Mantik bridges can be configured, e.g. by choosing model classes and parameters.
- The communication between bridges is handled via the Mantik Node Protocol (MNP).
- We provide adapters for data formats needed in the supported ML frameworks.
- Adaptation between data formats is handled automatically by the engine.



The KI:STE platform

Easing and promoting ML and AI usage in earth system science

Reproducible experiments via declarative workflow definition

- We provide curated and version controlled data in our **data projects**.
- All execution steps are version controlled.
- The workflow execution is **transparent** since all software is released in Open Source projects.
- Workflows can be defined in a declarative yaml format.
- Data references are included in the workflow definition to ensure **reproducibility**.
- **Collaboration** is easy. Just share your workflow definitions.



The KI:STE platform

Easing and promoting ML and AI usage in earth system science

Partners

Education

- Forschungszentrum Jülich
- RWTH Aachen
- University of Bonn
- University of Cologne

Commercial

- Ambrosys GmbH
- Mantik UG
- 52°North GmbH
- MAELSTROM