DEEPHEALTH

Deep-Learning and HPC to Boost Biomedical Applications for Health

DeepHealth Needs & Requirements for Benchmarking

Jon Ander Gómez (Technical Manager) - jon@upv.es

DataBench BenchLearning webinar

July 8th 2020

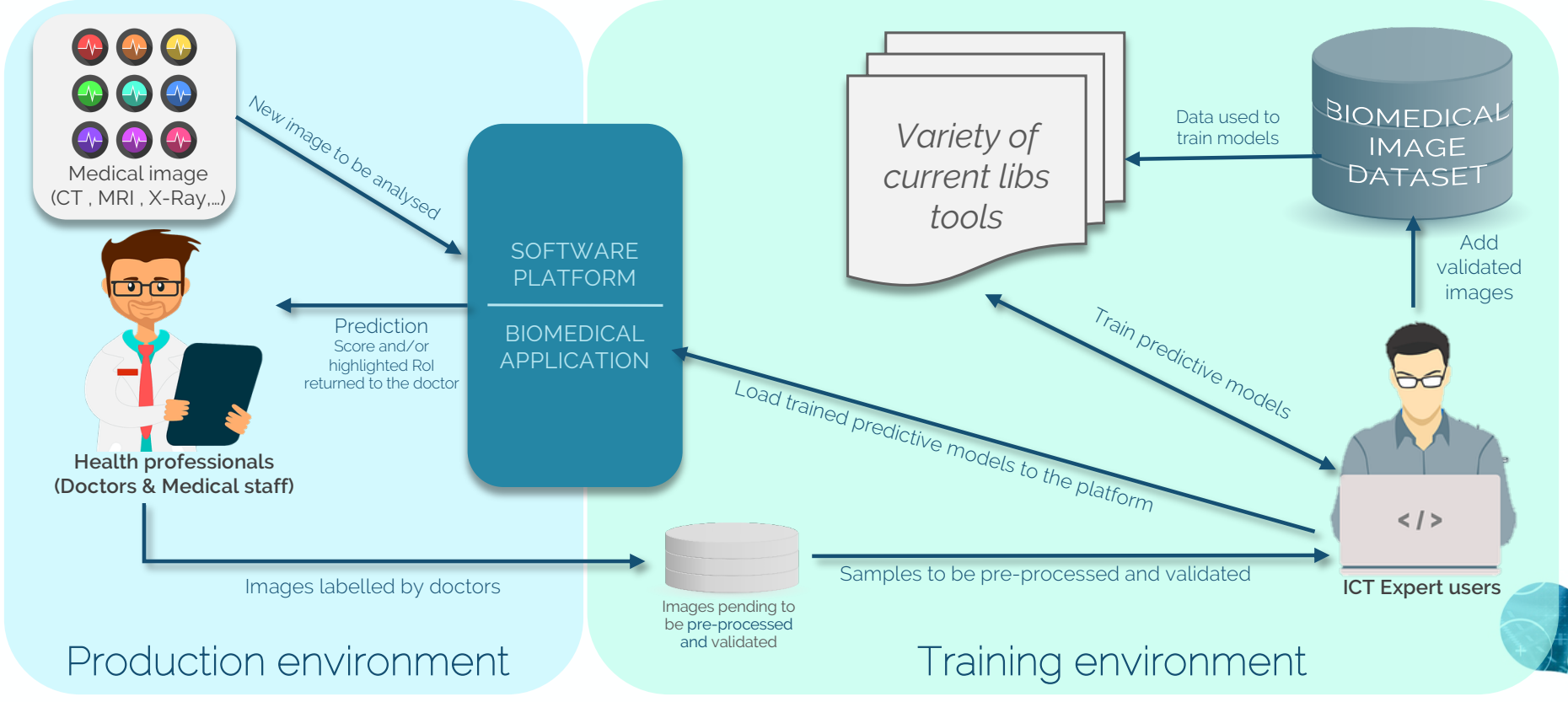
The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

A Little bit of context

- Healthcare: key sector in the global economy
- Public health systems generate large datasets of biomedical images
  - Large unexploited knowledge database
  - Interpretation of the clinical expert manually
- R & D on applying Artificial Intelligence (AI) to analyze biomedical images but . . .
  - Need for advanced skills in AI and different technologies and tools
  - Expensive processes in time and resources
  - Needs of high-quality data and take care of ethics
- HPC and BigData technologies (Big Data, HPC) sufficiently mature and available.

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

The scenario



The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

About DeepHealth

DEEPHEALTH

Aim & Goals

- **Put HPC computing power at the service of biomedical applications** with DL needs and apply DL techniques on large and complex image biomedical datasets to support **new and more efficient ways of diagnosis, monitoring and treatment of diseases.**
- Facilitate the daily work and **increase the productivity of medical personnel and IT professionals** in terms of image processing and the **use and training of predictive models** without the need of combining numerous tools.
- Offer a **unified framework** adapted to exploit underlying heterogeneous **HPC and Cloud architectures** supporting state-of-the-art and next-generation **Deep Learning (AI) and Computer Vision algorithms** to enhance European-based medical software platforms.

Key facts

Duration: 36 months
Starting date: Jan 2019

Budget 14.642.366 €
EU funding 12.774.824 €

22 partners from **9 countries**:
Research centers, Health organizations, large industries and SMEs

Research Organisations

Universitat Politècnica de València · Barcelona Supercomputing Center · list · UNIMORE · Karolinska Institutet · CRS4 · Università degli Studi di Torino · Otto von Guericke Universität Magdeburg

Health Organisations

Azienda Ospedaliero-Universitaria Città della Salute e della Scienza di Torino · Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana · CHUV · Spitalul Clinic Prof. Dr. Theodor Burghele București · Stockholms läns landsting

Large Industries

everis an NTT DATA Company · PHILIPS · THALES · SIMAVI Software Imagination & Vision

SMEs

WINGS ICT Solutions · TREE Technology · prodesign electronic gmbh · security Stelar research

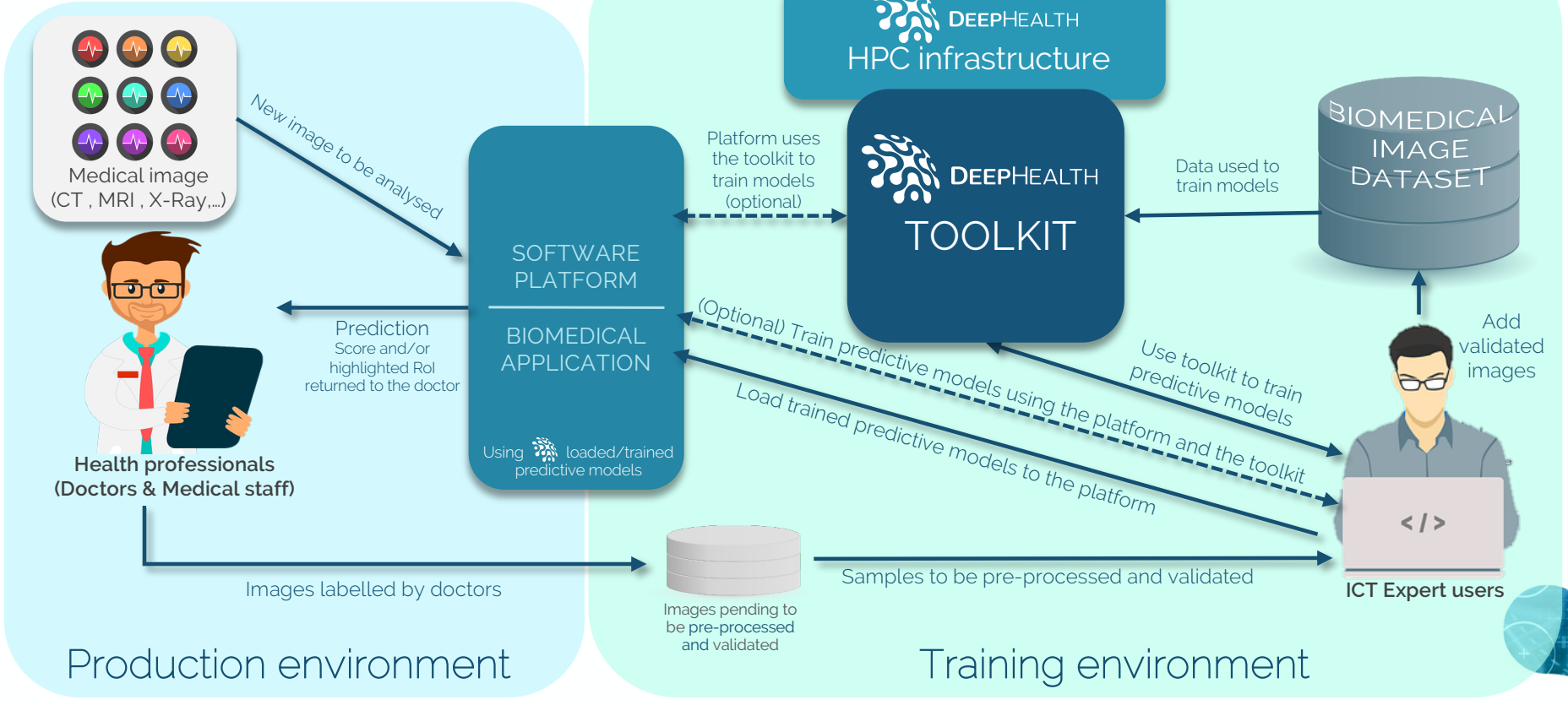
The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

Developments & Expected Results

- **The DeepHealth toolkit**
  - Free and open-source software: 2 libraries + front-end.
    - ***EDDLL:*** *The European Distributed Deep Learning Library*
    - ***ECVL:*** *the European Computer Vision Library*
  - Ready to run algorithms on Hybrid HPC + Cloud architectures with heterogeneous hardware (Distributed versions of the training algorithms)
  - Ready to be integrated into end-user software platforms or applications
- **HPC infrastructure** for an efficient execution of the training algorithms which are computational intensive by making use of heterogeneous hardware in a transparent way
- Seven enhanced **biomedical and AI software platforms** provided by EVERIS, PHILIPS, THALES, UNITO, WINGS, CRS4 and CEA that integrate the DeepHealth libraries to improve their potential
- Proposal for a structure for anonymised and pseudonymised data lakes
- **Validation** in 14 use cases (*Neurological diseases, Tumor detection and early cancer prediction, Digital pathology and automated image annotation*).

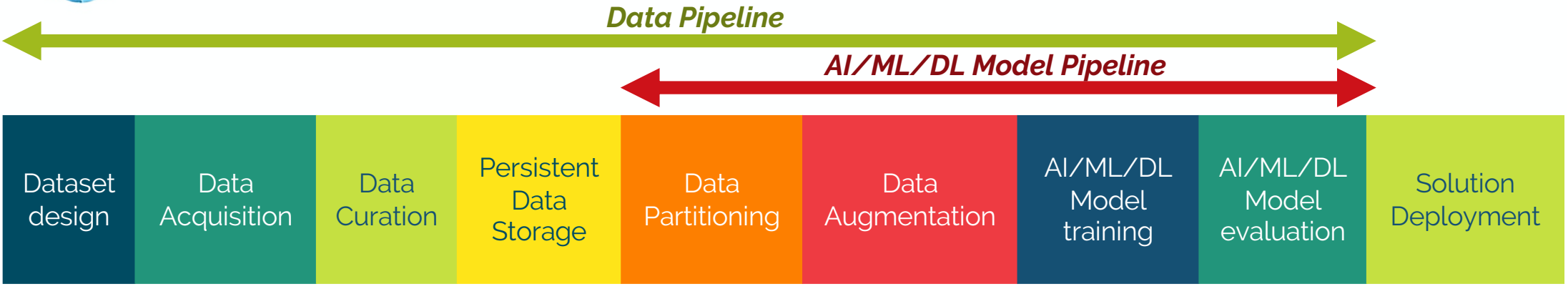
The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

The Concept



The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

Data & Model Pipelines



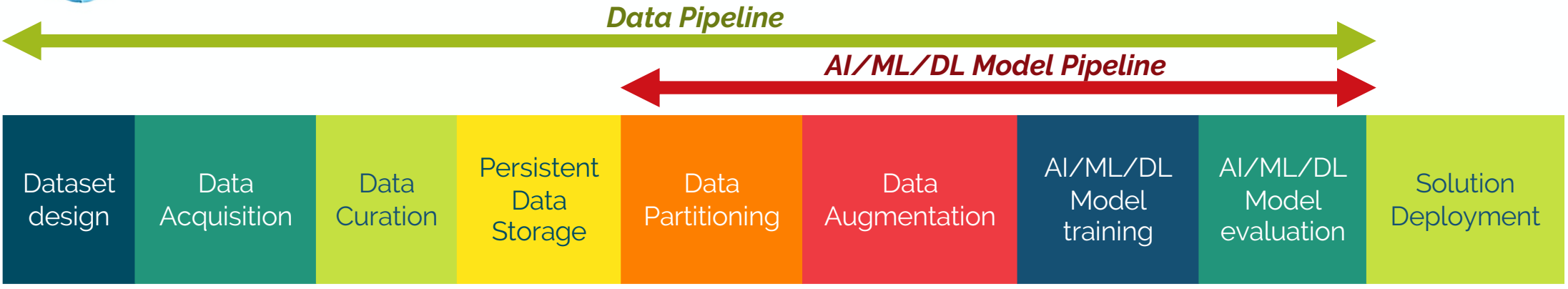
Model Pipeline is considered part of the ***Data Pipeline***

Both pipelines are suitable for business and research applications

The whole ***Data Pipeline*** is applicable to any sector. Our project is focused on the Health sector

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

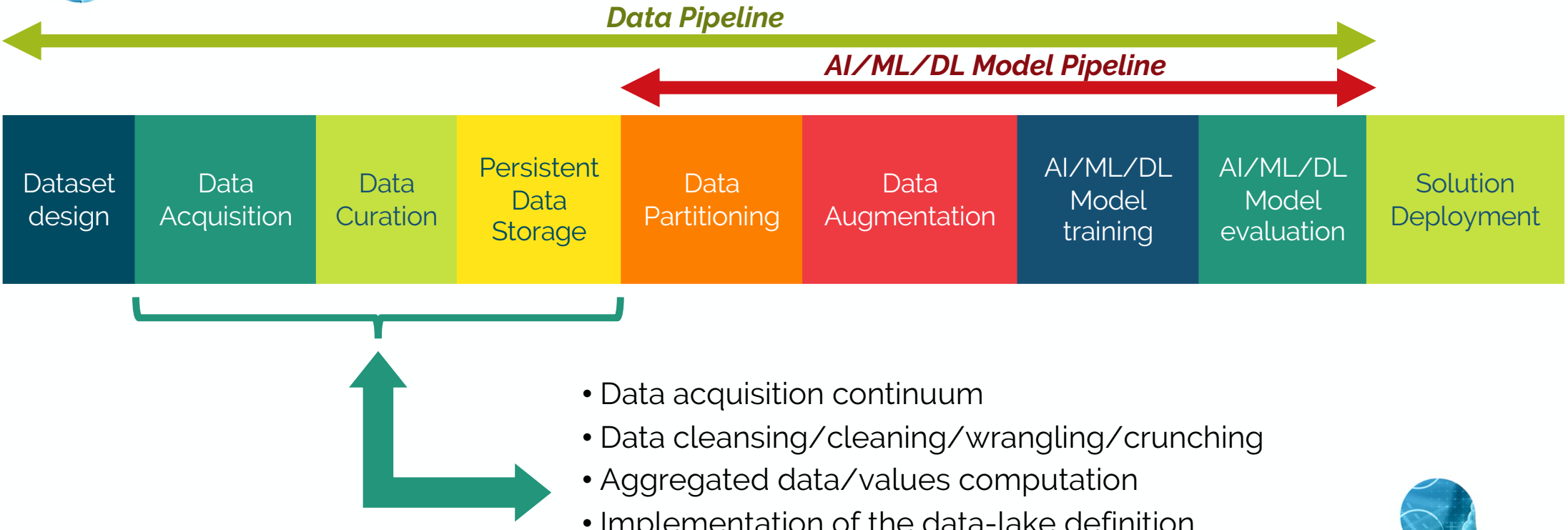
Data & Model Pipelines – Dataset Design



- Data types and formats –identification and definition
- Metadata definition
- Data Lake structure definition
- Guidelines / HOW-TOs
- Etc.

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

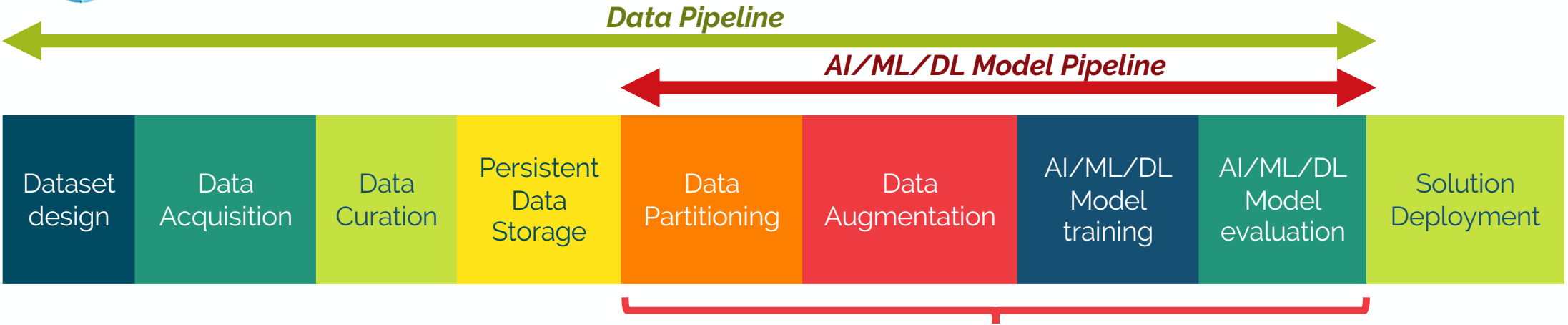
Data & Model Pipelines – Acquisition + Curation + Storage



- Data acquisition continuum
- Data cleansing/cleaning/wrangling/crunching
- Aggregated data/values computation
- Implementation of the data-lake definition
- Creation of users and permissions or make data public

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

Data & Model Pipelines – Model Pipeline



Model training loop:

- Partitioning in training/validation/test subsets
- Data augmentation on-the-fly
- Training and evaluation of models
- Cloud & High Performance Computing requirements

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

Needs & Requirements

Evaluate **datasets** in terms of

1. **Findability** – where should a data scientist search for the dataset?
2. **Availability** – how long does a data scientist need to start the initial exploratory data analysis?
3. **Interoperability** – how long does a data scientist need to start training AI/ML/DL models with a dataset?
4. **Reusability** – are previously obtained results with a dataset public and available to other researchers / data scientists?
5. **Privacy / Anonymisation** – can the dataset be made public without compromising the identity of individuals?
6. **Quality** – is the dataset biased or unbalanced? What procedure has been followed to validate annotations?

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

Needs & Requirements

Evaluate **Deep Learning libraries** in terms of

1. **Speed-up** – is distributed learning really efficient?
2. **Convergence** – does the distributed learning reach the same model accuracy in less time?
3. **Usability** – how long does a developer need to use the libraries effectively?
4. **Integrability** – how difficult is it to integrate the libraries as part of solutions to deploy?
5. **KPI**s: time-of-training-models (**totm**), performance/power/accuracy trade-off, etc.
6. **Others** – can you help us to evaluate other aspects?

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

Needs & Requirements

Evaluate **Software Platforms** in terms of

1. **Usability** – how long does a domain application expert need to manage the software tool effectively?
2. **Completeness** – does the application platform provide all the algorithms/procedures/functions to allow domain application experts to easily define the sequences of steps to implement the data and/or model pipelines?
3. **Compatibility** – how many data formats does the platform admits to import/export data and models from/to other frameworks?
4. **KPI**s: time-to-model-in-production (**ttmip**), time-of-pre-processing-images (**toppi**), etc.
5. **Others** – can you help us to evaluate other aspects?

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.

DEEPHEALTH

Questions?

Technical Manager: Jon Ander Gómez jon@upv.es

https://deephealth-project.eu

@DeepHealthEU

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825111.