EDIHs and AI-on-demand platform: getting to know each other

Moderated by Yves Paindaveine, DG CNECT

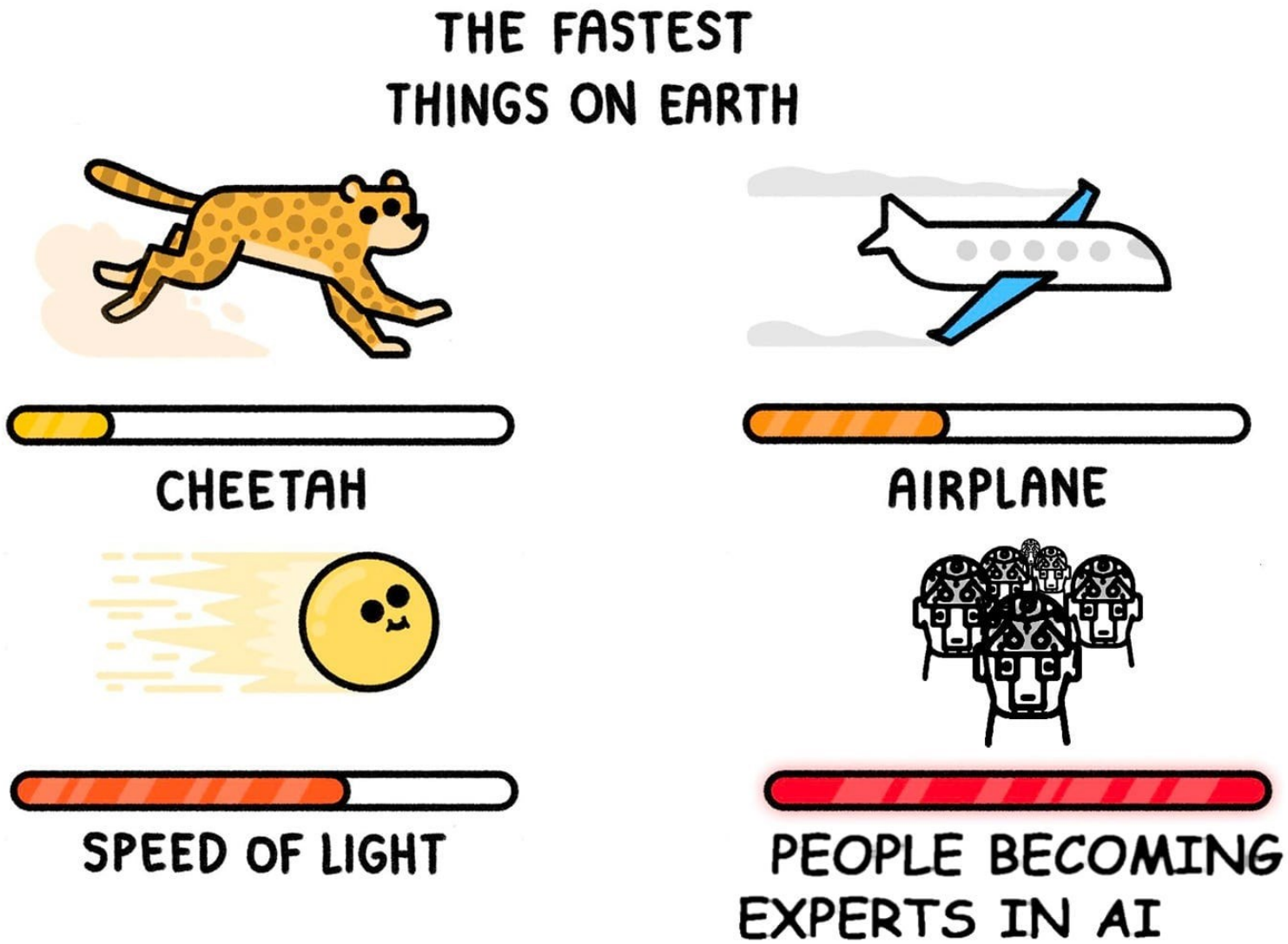
European Commission

Some general points

The meeting will be recorded and made available online.

Please keep your microphone muted when not speaking.

Please raise your hand when you wish to speak. Unmute your mic when the chair invites you to take the floor.



Outline of the webinar

1. Presentation by David Zuñel, DG CONNECT/A1
2. PART I : The global picture, a brief overview of the AI-on-Demand Platform
3. PART II : Hands-on session for practitioners
4. Discussion & Conclusion

   (we will then open the microphones & video for all)

AI on Demand platform EDIHs Webinar

24th April, 2024

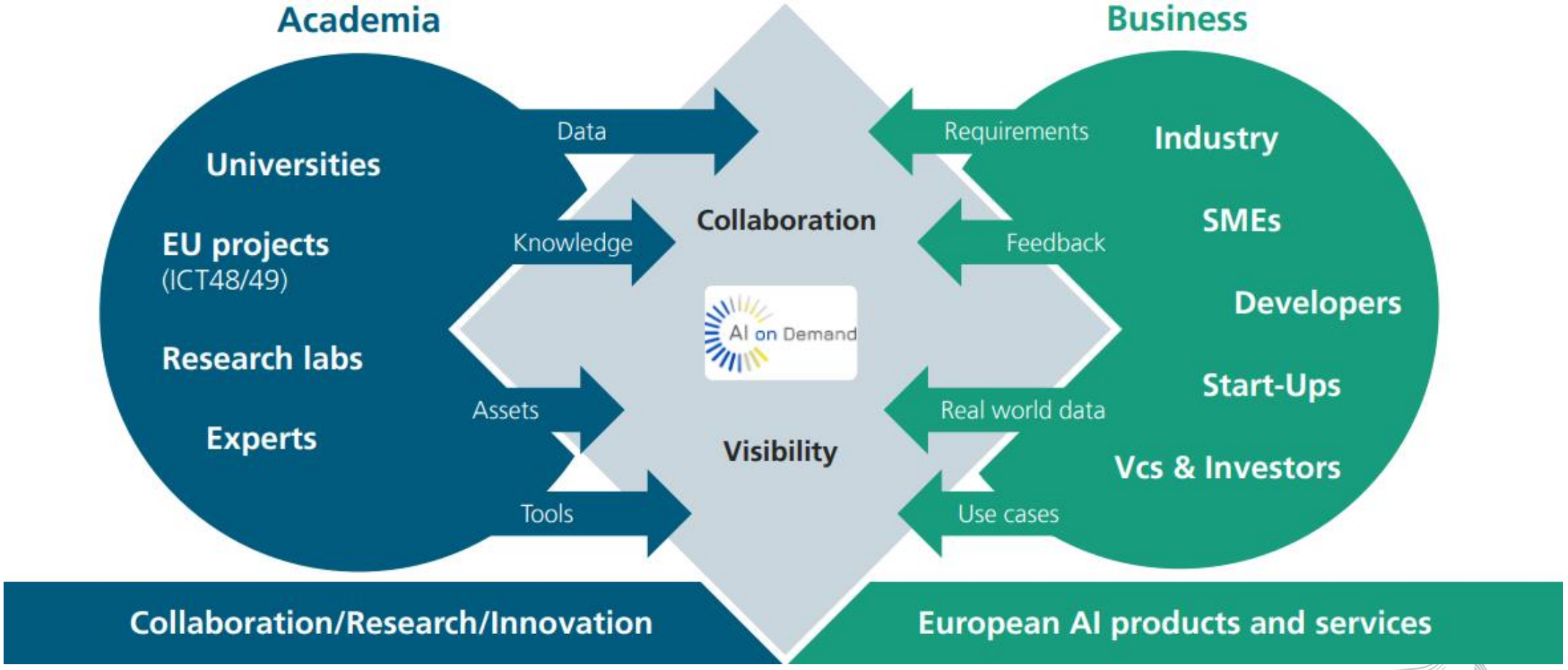
David Zuñel Ballester
Robotics & Artificial Intelligence – CNECT A1
European Commission

Agenda

PART I : The global picture			
Intro	How the AIoD fits in the overall EU AI ecosystem	5’	European Commission
AIoD overview	Description on the AIoD platform. Current services, short-term and future outlook. AIoD Branding & Communication	35’	University College Cork (AI4Europe project)
DeployAI	Future outlook	15’	Fraunhofer Demokritos Aalto University (DeployAI project)
PART II : Hands-on session for practitioners			
Platform walk-through	Metadata catalogue. Integrating a new service + RAIL. AI Builder.	50’	University College Cork Fraunhofer (AI4Europe project)
Q&A		15’	All

The European AI on Demand platform (AIoD)

A bridge & Catalyst between European AI research, industry and public services

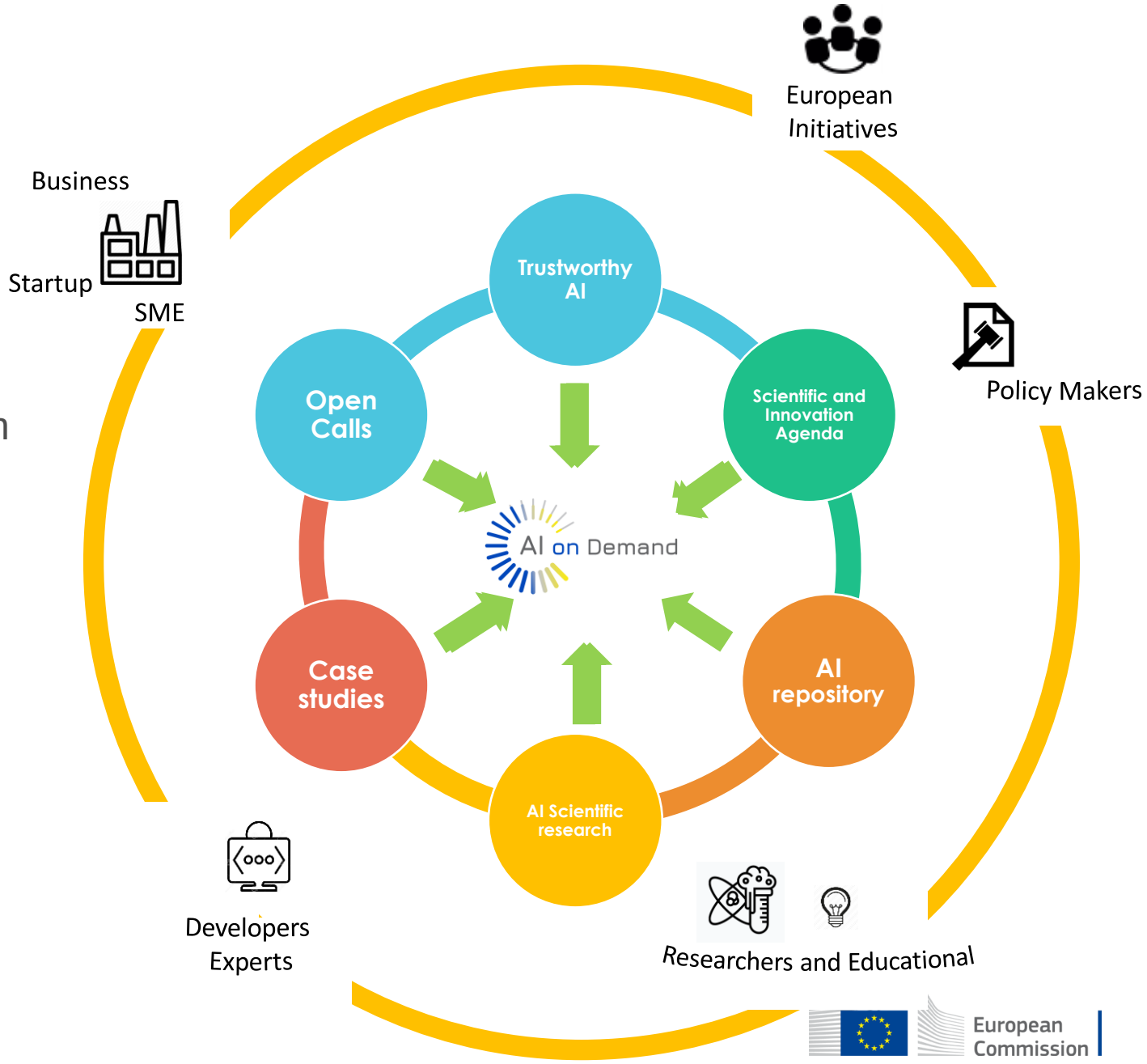


AIoD objectives

Trustworthy AI made in Europe

- Avoid the fragmentation of the European AI landscape
- One-stop-shop for trustworthy AI made in Europe
- Lower the barriers for European industry and public services to use and access AI technologies
- Enhance sharing, collaboration and networking across AI stakeholders
- Support ethical and regulatory European pillars

https://aiod.eu/

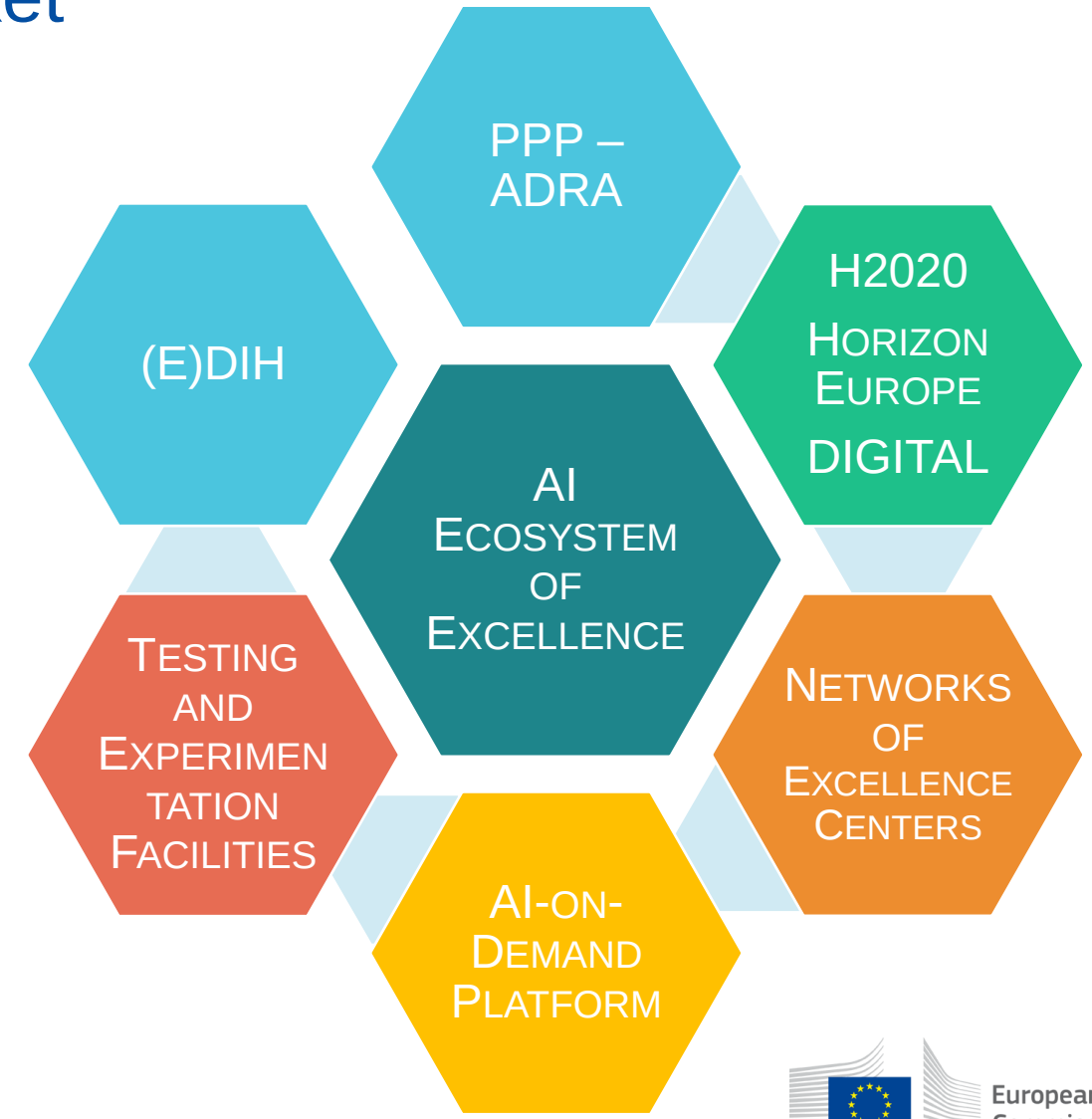


European Commission

Ecosystem of Excellence in Robotics & AI

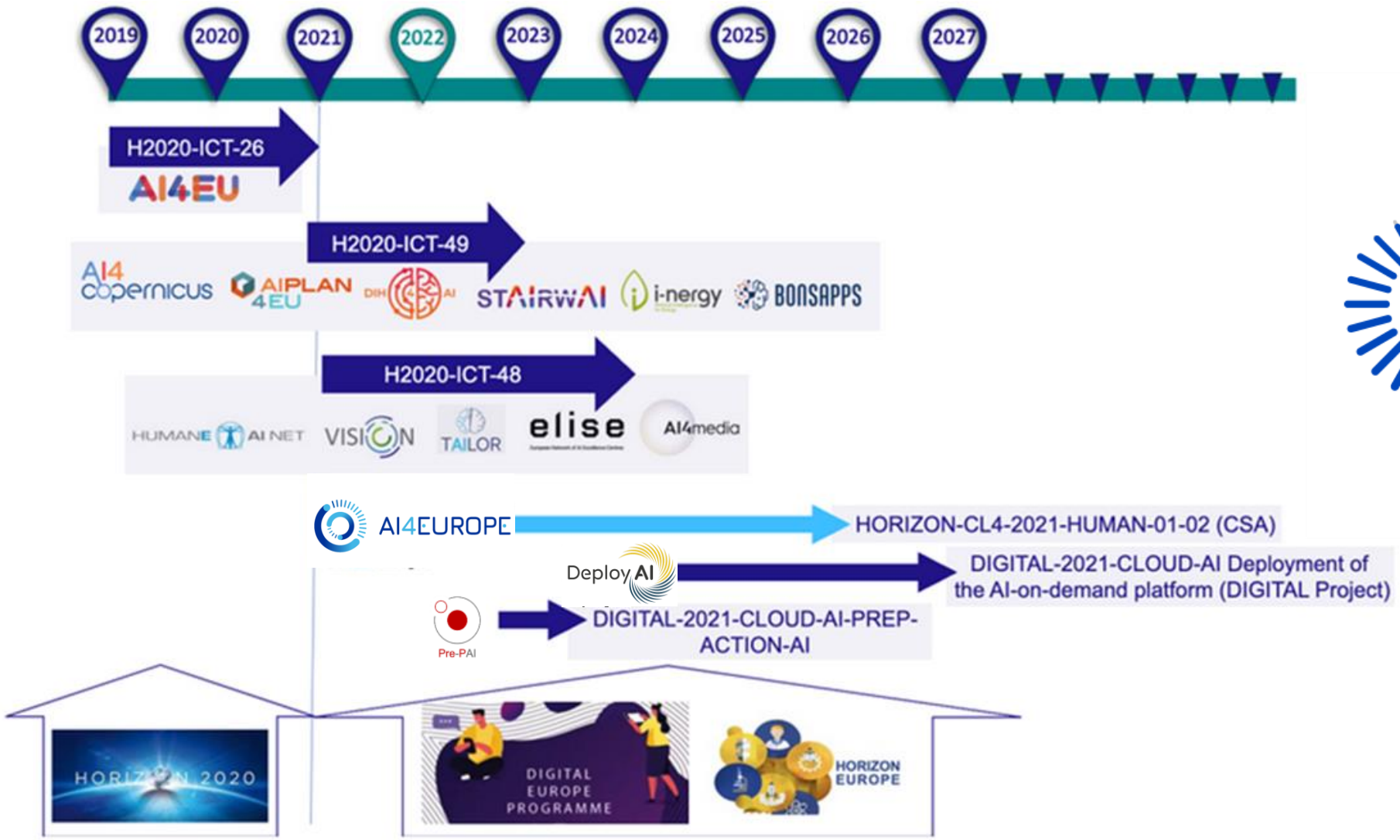
Support from the lab, to the market

- Collaboration between projects, projects WPs and with the CSAs (Adra-e, AI4Europe, DeployAI)
- Investment: 1Bn€/Year EU FUNDING → 20Bn€/Year invest @EU level
- Ambition: by 2030, 75% of European enterprises have taken up AI



European Commission

AIoD roadmap

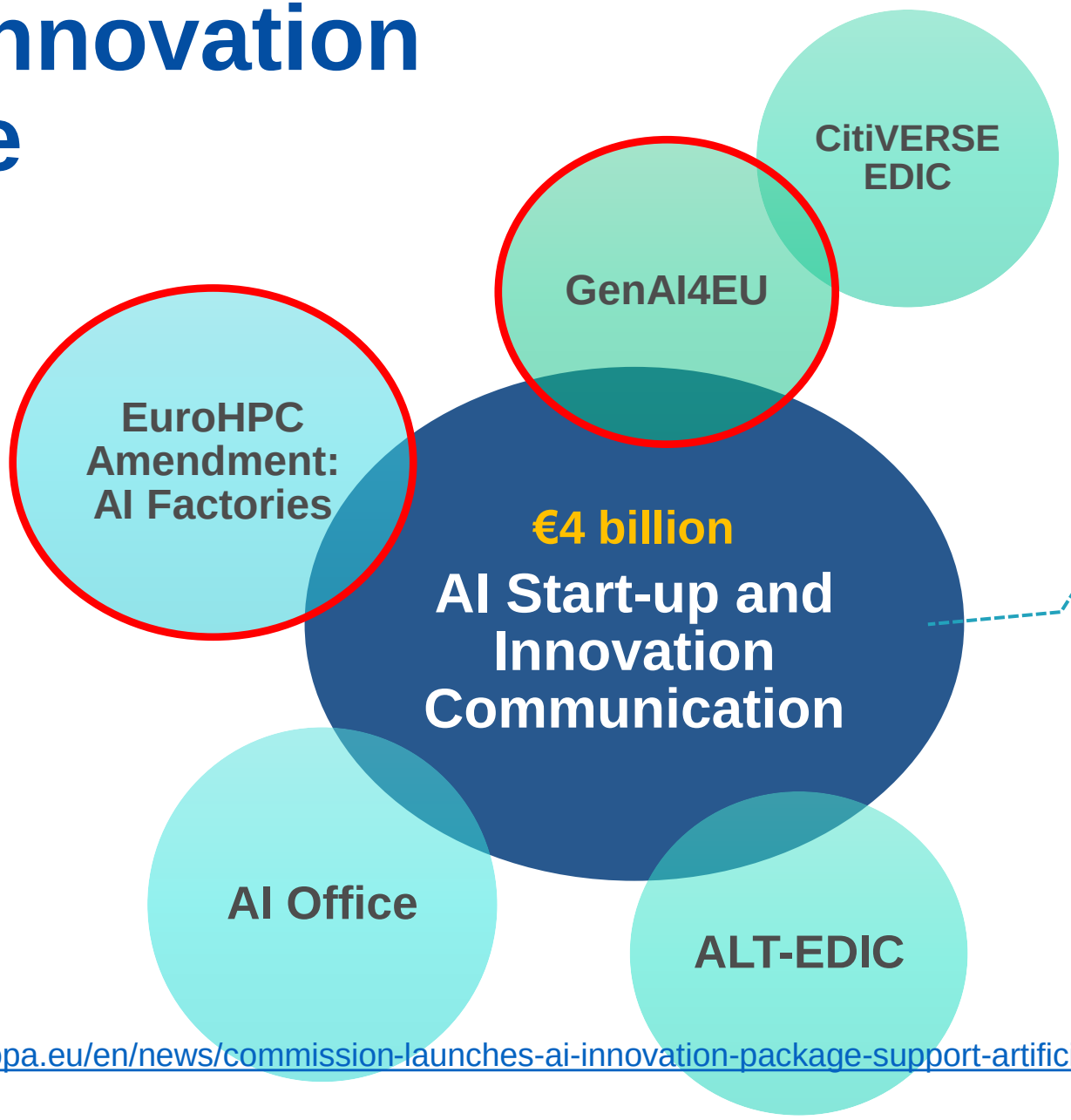


AI on Demand

AI on Demand

European Commission

The AI Innovation Package



Key Ingredients for AI

- Data Access
- Algorithms
- Investment
- Skills
- Processors

https://digital-strategy.ec.europa.eu/en/news/commission-launches-ai-innovation-package-support-artificial-intelligence-startups-and-smes

Thank you very much for your attention.

Enjoy the webinar!

European Commission

AI on Demand

Knowledge and services for the AI community

AI-on-demand

Current services and developments

EDIHs webinar

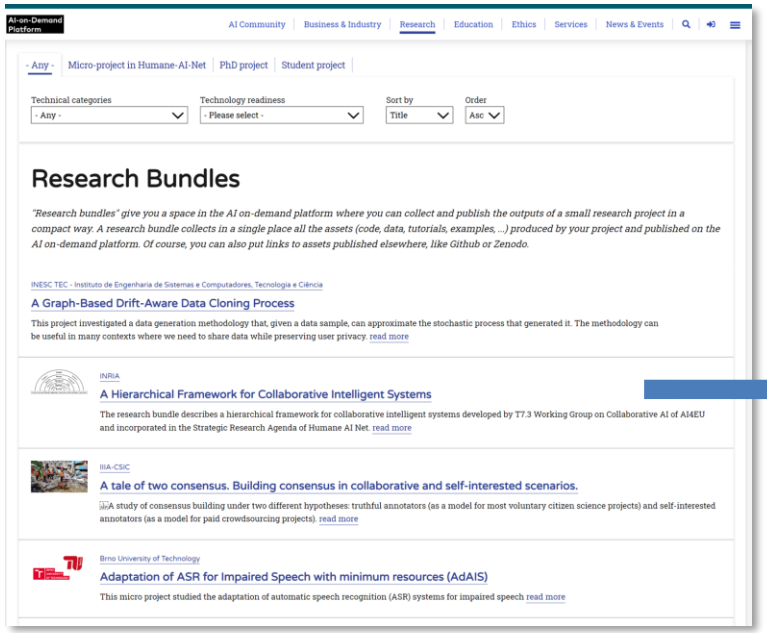
24-Apr-2024

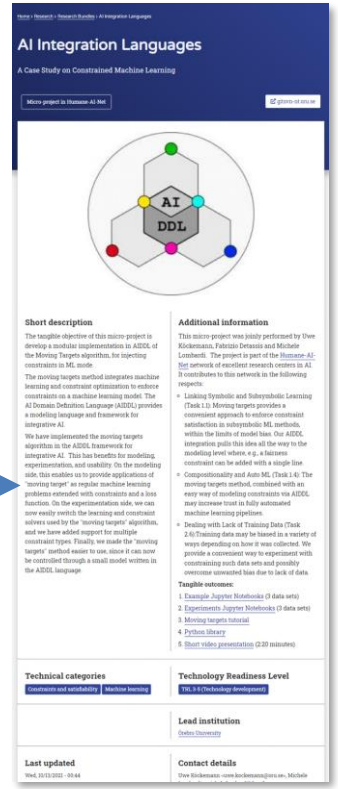
Prof. Barry O'Sullivan

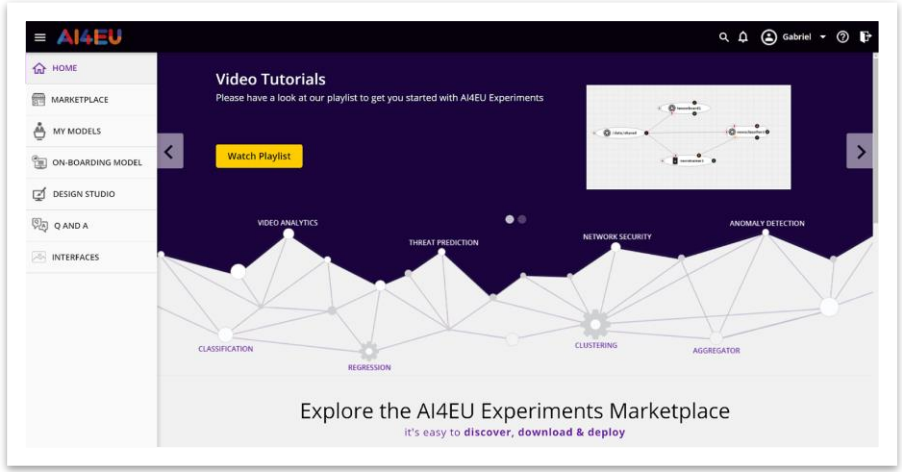
University College Cork (UCC)

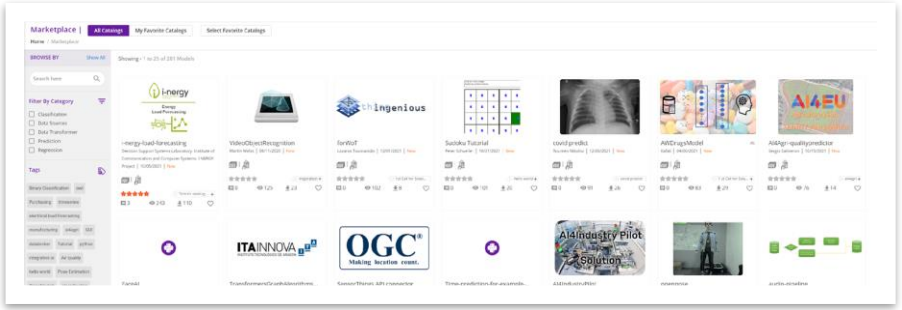
Funded by the European Union

AI4EU former systems - 2020









1. A portal – catalogues (DRUPAL)

2. Tool for composition of containers – (ACUMOS/Eclipse Graphene/AI4Experiments)

AI4EU former systems - 2020

1. Dev environment: Drupal
2. Operational (391 AI assets)
3. Driven by AI4Europe
4. Several catalogues – Research bundles, news, success stories, ..

1. **Constrained to integrate**
2. **Hard to find Drupal developers**
3. **Force the developers to develop in Drupal**

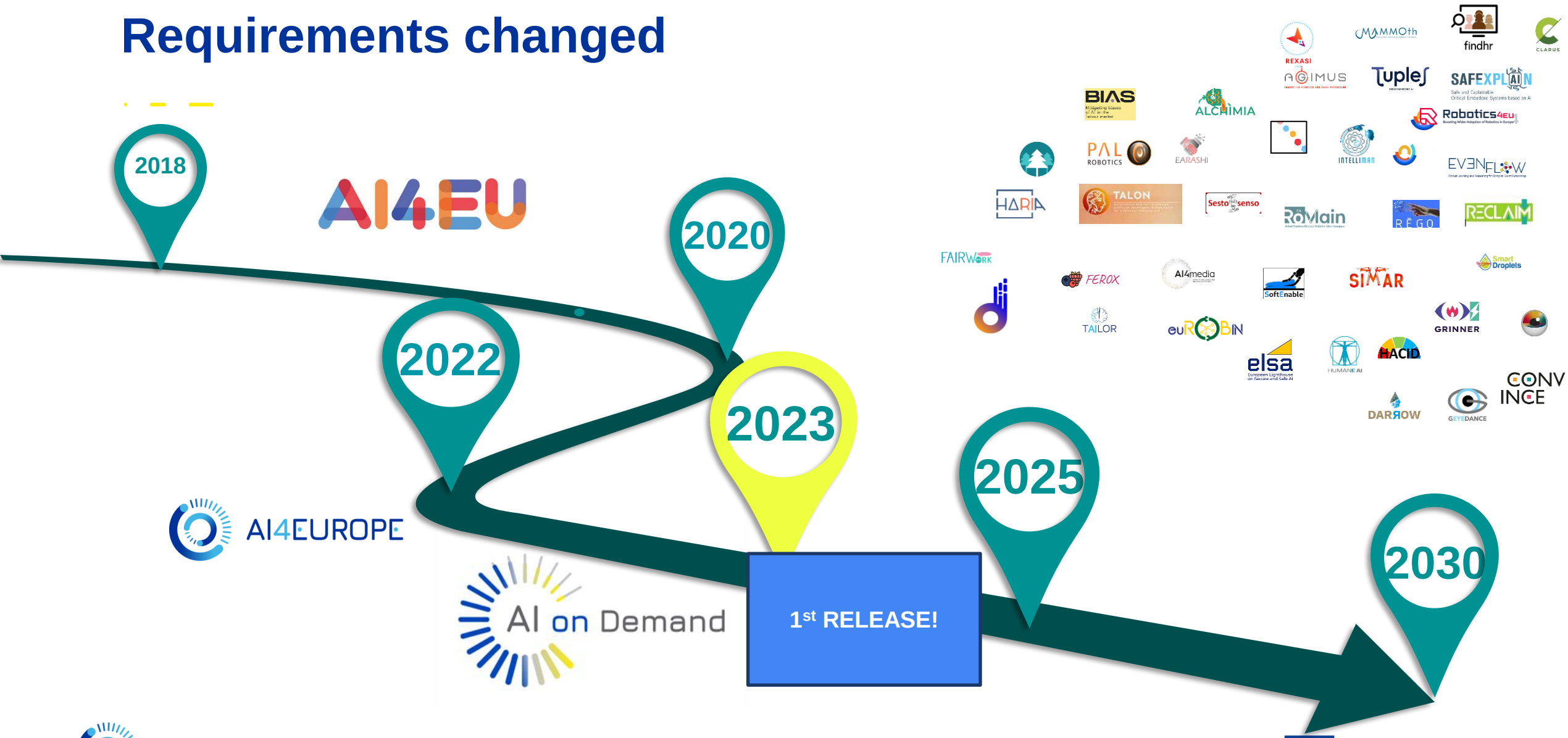
1. The portal – catalogues (DRUPAL)

1. Linux distribution (Eclipse Graphene)
2. Operational (329 Models)
3. Allows composition of pipelines
4. AI Playground for deployments

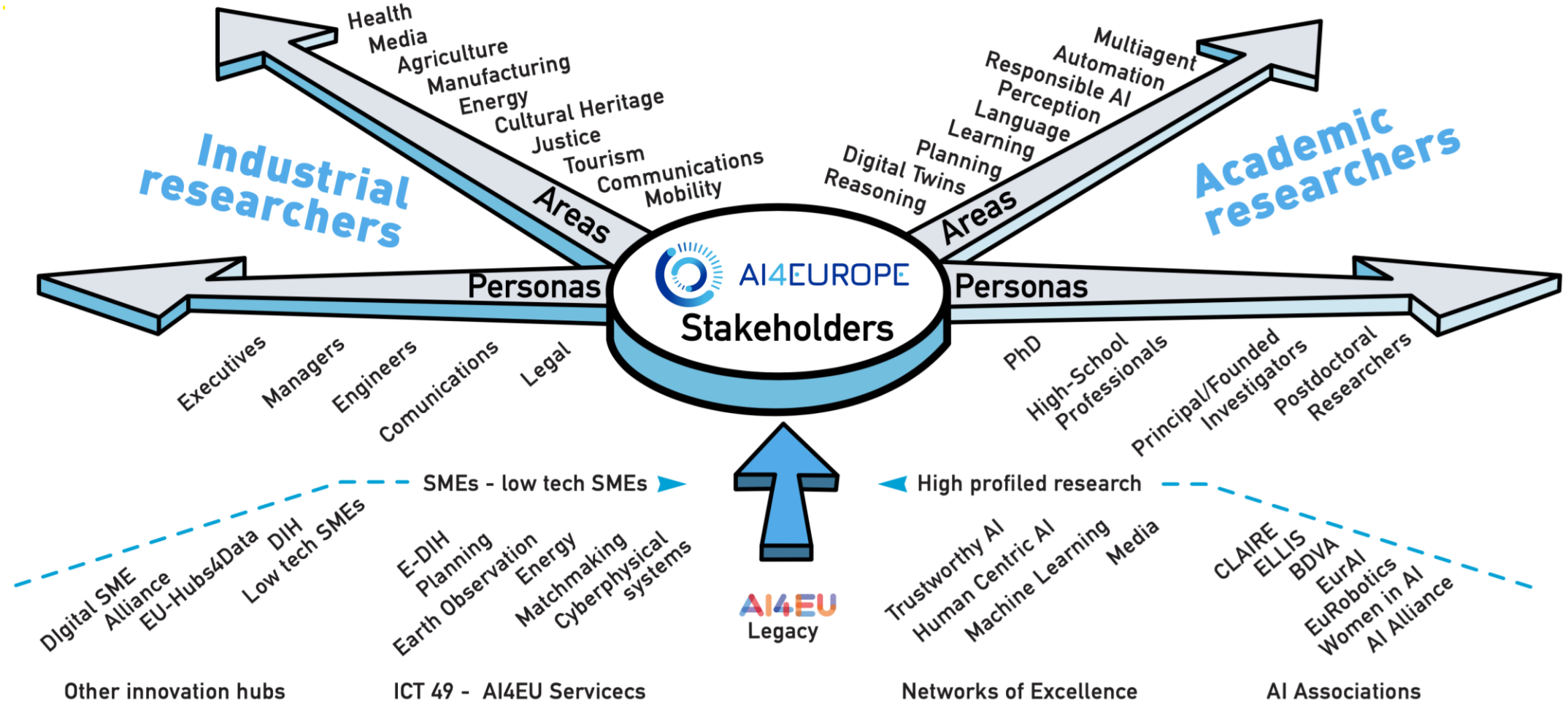
1. **Force the developers to understand the whole layered system**
2. **Some constraints on components to be incorporated**

2. Tool for composition of containers – (ACUMOS/Eclipse Graphene/AI4Experiments)

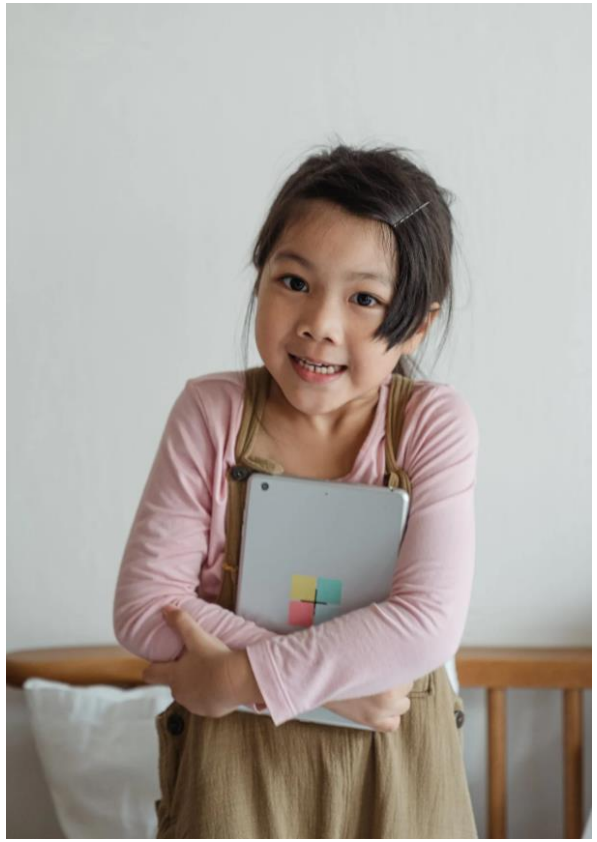
Requirements changed



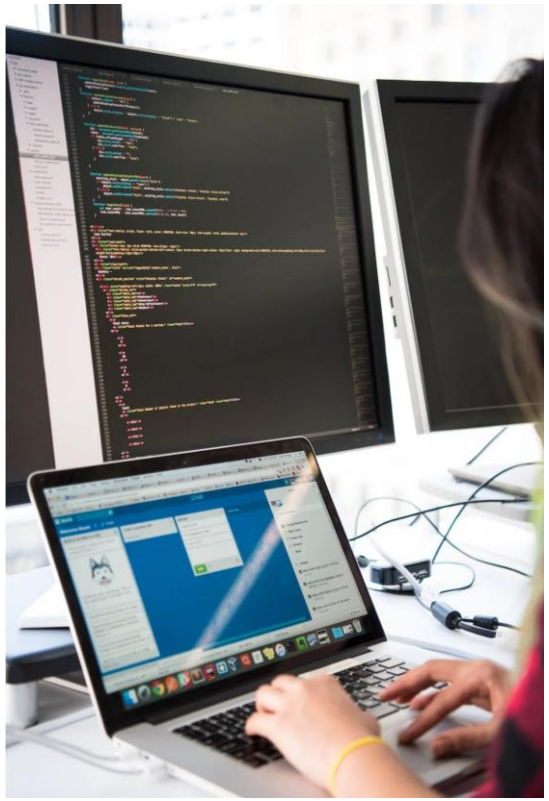
Potential Users



Simplified potential users



AI Consumers & Service Developers



What does the Consumers need?

Big Iron, Big Data, and Big Identity

Craig A. LEE [a,1], Marcio ASSIS [b], Luiz F. BITTENCOURT [b], Stefano NATIVI [c] and Rafael TOLOSANA-CALASANZ [d]
[a] *The Aerospace Corporation*
[b] *University of Campinas*
[c] *National Research Council of Italy*
[d] *University of Zaragoza*

Abstract. While High-Performance Computing (HPC) typically focuses on very large, parallel machines, i.e., Big Iron, running massive numerical codes, the importance of extracting knowledge from massive amounts of information, i.e., Big Data, has been clearly recognized. While many massive data sets can be produced within a single administrative domain, many more massive data sets can be, and must be, assembled from multiple sources. Aggregating data from multiple sources can be a tedious task. First, the locations of the desired data must be known. Second, access to the data sets must be allowed. For publicly accessible data, this may not pose a serious problem. However, many application domains and user groups may wish to facilitate, and have some degree of control over, how their resources are discovered and shared. Such collaboration requirements are addressed by federation management technologies. In this paper, we argue that effective, widely-adopted federation management tools, i.e., Big Identity, are critical for enabling many Big Data applications, and will be central to how the Internet of Things is managed. To this end, we re-visit the NIST cloud deployment models to extract and identify the fundamental aspects of federation management: crossing trust boundaries, trust topologies, and deployment topologies. We then review possible barriers to adoption and relevant, existing tooling and standards to facilitate the emergence of a common practice for Big Identity.

Keywords. big data, identity, federation management, deployment models

1. Introduction

The need to share data, and computing resources in general, is fundamental. This need has driven the development of computing networks and the World Wide Web. All segments of society – academia, arts, business and government – increasingly rely on electronic communication. All of this communication and the devices involved are, in fact, converging into an *Internet of Things (IoT)*.

[1] Corresponding Author: The Aerospace Corporation M1-102, 2310 East El Segundo Blvd., El Segundo, CA 90245-4691, USA, E-Mail: lee@aero.org.

Big Iron

Big Data

Big Identity

[1] Lee Craig, Marcio ASSIS, Luiz BITTENCOURT, Stefano NATIVI, Rafael TOLOSANA-CALASANZ
New Frontiers in High Performance Computing and Big Data, 2017
Publication year: 2017

What does the Consumers need?

Big Iron, Big Data, and Big Identity

Craig A. LEE [a,1], Marcio ASSIS [b], Luiz F. BITTENCOURT [b], Stefano NATIVI [c] and Rafael TOLOSANA-CALASANZ [d]
[a] *The Aerospace Corporation*
[b] *University of Campinas*
[c] *National Research Council of Italy*
[d] *University of Zaragoza*

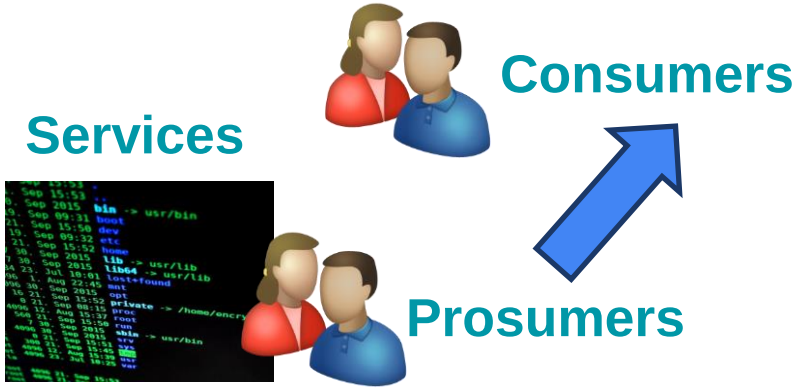
Abstract. While High-Performance Computing (HPC) typically focuses on very large, parallel machines, i.e., Big Iron, running massive numerical codes, the importance of extracting knowledge from massive amounts of information, i.e., Big Data, has been clearly recognized. While many massive data sets can be produced within a single administrative domain, many more massive data sets can be, and must be, assembled from multiple sources. Aggregating data from multiple sources can be a tedious task. First, the locations of the desired data must be known. Second, access to the data sets must be allowed. For publicly accessible data, this may not pose a serious problem. However, many application domains and user groups may wish to facilitate, and have some degree of control over, how their resources are discovered and shared. Such collaboration requirements are addressed by federation management technologies. In this paper, we argue that effective, widely-adopted federation management tools, i.e., Big Identity, are critical for enabling many Big Data applications, and will be central to how the Internet of Things is managed. To this end, we re-visit the NIST cloud deployment models to extract and identify the fundamental aspects of federation management: crossing trust boundaries, trust topologies, and deployment topologies. We then review possible barriers to adoption and relevant, existing tooling and standards to facilitate the emergence of a common practice for Big Identity.

Keywords. big data, identity, federation management, deployment models

1. Introduction

The need to share data, and computing resources in general, is fundamental. This need has driven the development of computing networks and the World Wide Web. All segments of society – academia, arts, business and government – increasingly rely on electronic communication. All of this communication and the devices involved are, in fact, converging into an *Internet of Things (IoT)*.

[1] Corresponding Author: The Aerospace Corporation M1-102, 2310 East El Segundo Blvd., El Segundo, CA 90245-4691, USA, E-Mail: lee@aero.org.



Big Iron

Big Data

Big Identity

[1] Lee Craig, Marcio ASSIS, Luiz BITTENCOURT, Stefano NATIVI, Rafael TOLOSANA-CALASANZ
New Frontiers in High Performance Computing and Big Data, 2017
Publication year: 2017

What does the Consumers need?

Big Iron, Big Data, and Big Identity

Craig A. LEE [a,1], Marcio ASSIS [b], Luiz F. BITTENCOURT [b], Stefano NATIVI [c] and Rafael TOLOSANA-CALASANZ [d]
[a] *The Aerospace Corporation*
[b] *University of Campinas*
[c] *National Research Council of Italy*
[d] *University of Zaragoza*

Abstract. While High-Performance Computing (HPC) typically focuses on very large, parallel machines, i.e., Big Iron, running massive numerical codes, the importance of extracting knowledge from massive amounts of information, i.e., Big Data, has been clearly recognized. While many massive data sets can be produced within a single administrative domain, many more massive data sets can be, and must be, assembled from multiple sources. Aggregating data from multiple sources can be a tedious task. First, the locations of the desired data must be known. Second, access to the data sets must be allowed. For publicly accessible data, this may not pose a serious problem. However, many application domains and user groups may wish to facilitate, and have some degree of control over, how their resources are discovered and shared. Such collaboration requirements are addressed by federation management technologies. In this paper, we argue that effective, widely-adopted federation management tools, i.e., Big Identity, are critical for enabling many Big Data applications, and will be central to how the Internet of Things is managed. To this end, we re-visit the NIST cloud deployment models to extract and identify the fundamental aspects of federation management: crossing trust boundaries, trust topologies, and deployment topologies. We then review possible barriers to adoption and relevant, existing tooling and standards to facilitate the emergence of a common practice for Big Identity.

Keywords. big data, identity, federation management, deployment models

1. Introduction

The need to share data, and computing resources in general, is fundamental. This need has driven the development of computing networks and the World Wide Web. All segments of society – academia, arts, business and government – increasingly rely on electronic communication. All of this communication and the devices involved are, in fact, converging into an *Internet of Things (IoT)*.

[1] Corresponding Author: The Aerospace Corporation M1-102, 2310 East El Segundo Blvd., El Segundo, CA 90245-4691, USA, E-Mail: lee@aero.org.

Consumers

Services

Prosumers

Tools

Big Iron

Big Data

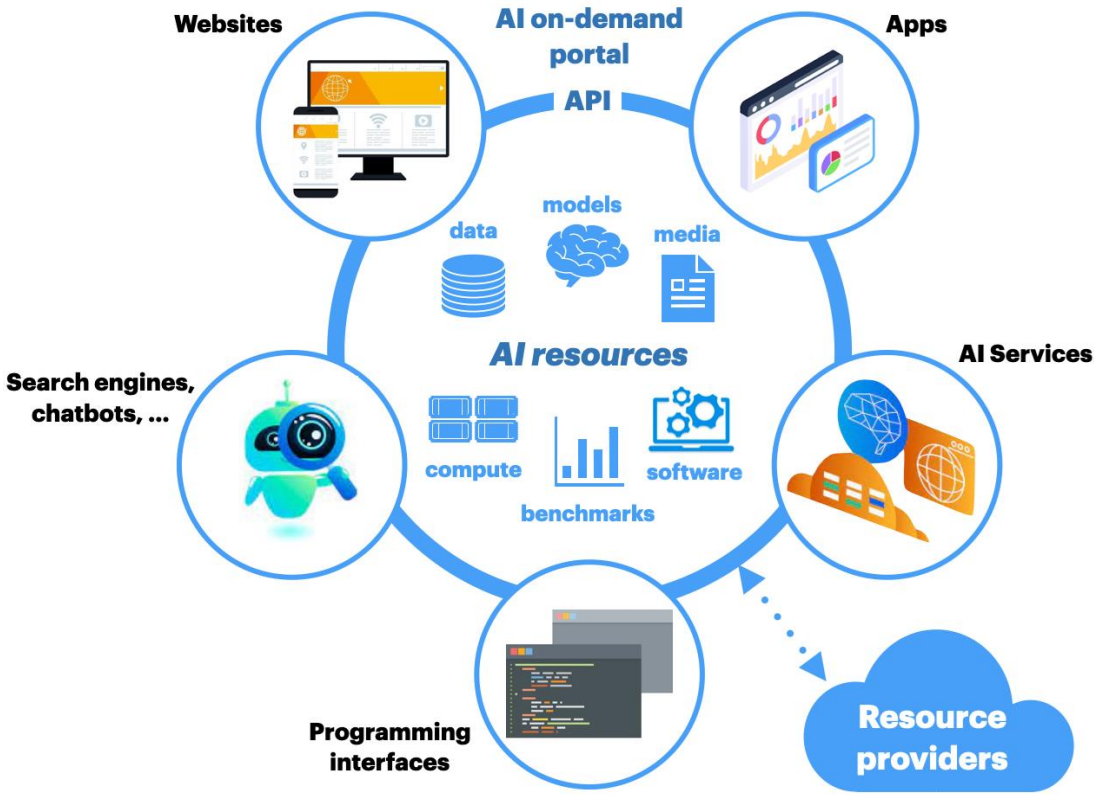
Big Identity

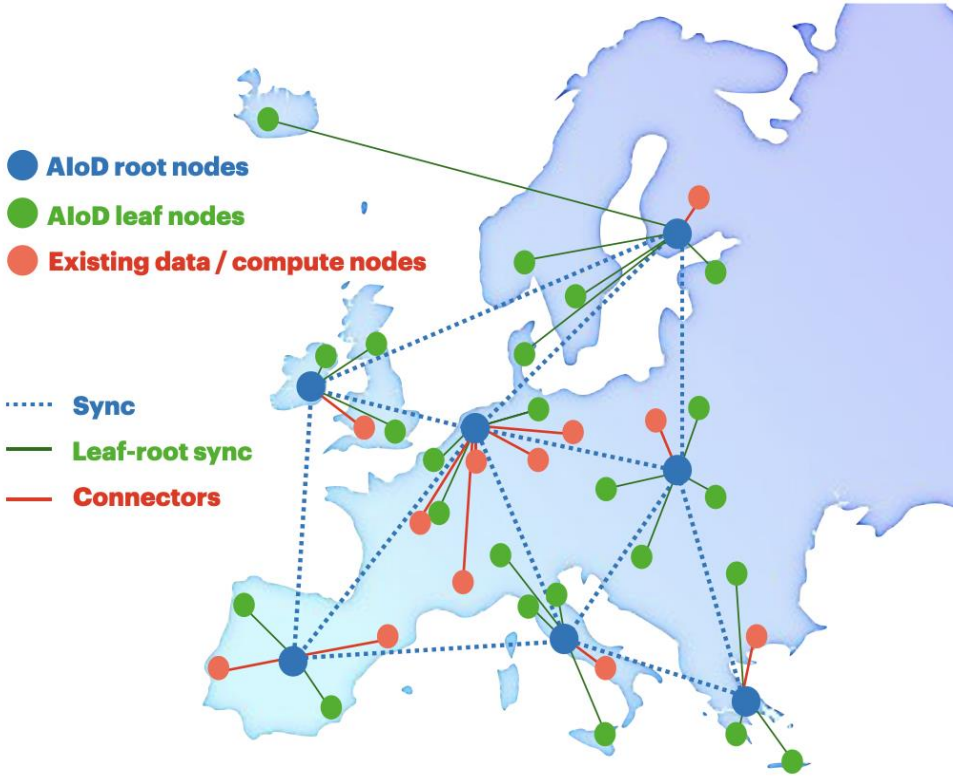
[1] Lee Craig, Marcio ASSIS, Luiz BITTENCOURT, Stefano NATIVI, Rafael TOLOSANA-CALASANZ
New Frontiers in High Performance Computing and Big Data, 2017
Publication year: 2017

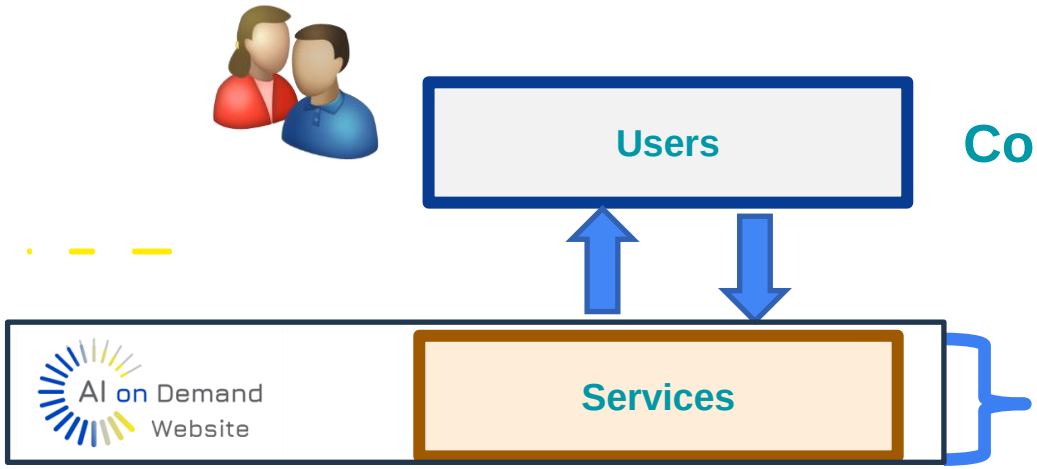
The new concept

AI4EUROPE

High level view of the AIoD platform



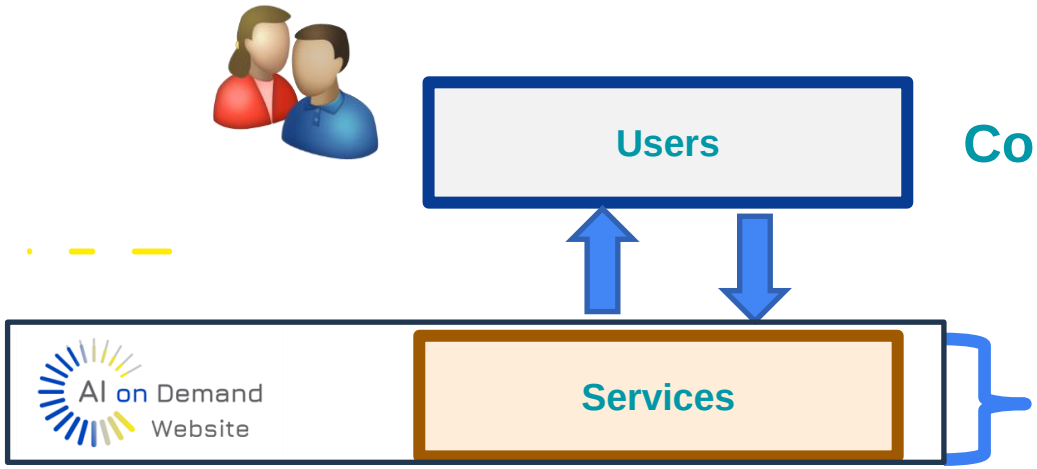




Consumers

Developers
Projects
Companies
DIHs

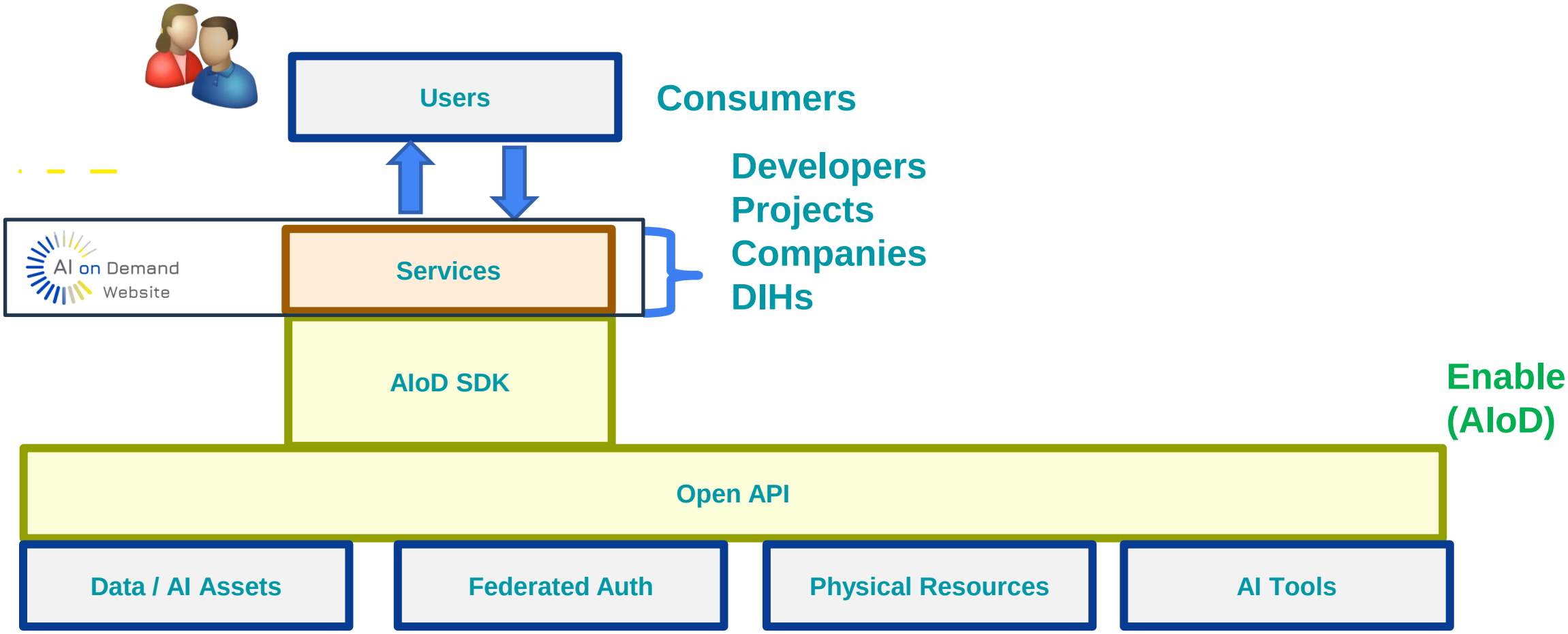
- To foster the creation of services.
- Not to focus on “catalogues only”.
- To provide the commonalities to support the consumption and production of AI



Consumers

Developers
Projects
Companies
DIHs

Data / AI Assets	Federated Auth	Physical Resources	AI Tools
IDS Companies Research Centres Multiple platforms		EOSC Companies Federated resources	AITANA Deep Hybrid Cloud …


Users
Consumers
Developers
Projects
Companies
DIHs
AI on Demand Website
Services
AIoD SDK
Enabler (AIoD)
Open API
Data / AI Assets
Federated Auth
Physical Resources
AI Tools
IDS
Companies
Research Centres
Multiple platforms
EOSC
Companies
Federated resources
AITANA
Deep Hybrid Cloud
…

Users

Consumers

Developers
Projects
Companies
DIHs

AI on Demand Website

Services

Next releases !
(More services)

AIoD SDK

Enabler
(AIoD)

Open API

Data / AI Assets

Federated Auth

Physical Resources

AI Tools

IDS
Companies
Research Centres
Multiple platforms

EOSC
Companies
Federated resources

AITANA
Deep Hybrid Cloud
… Next releases !

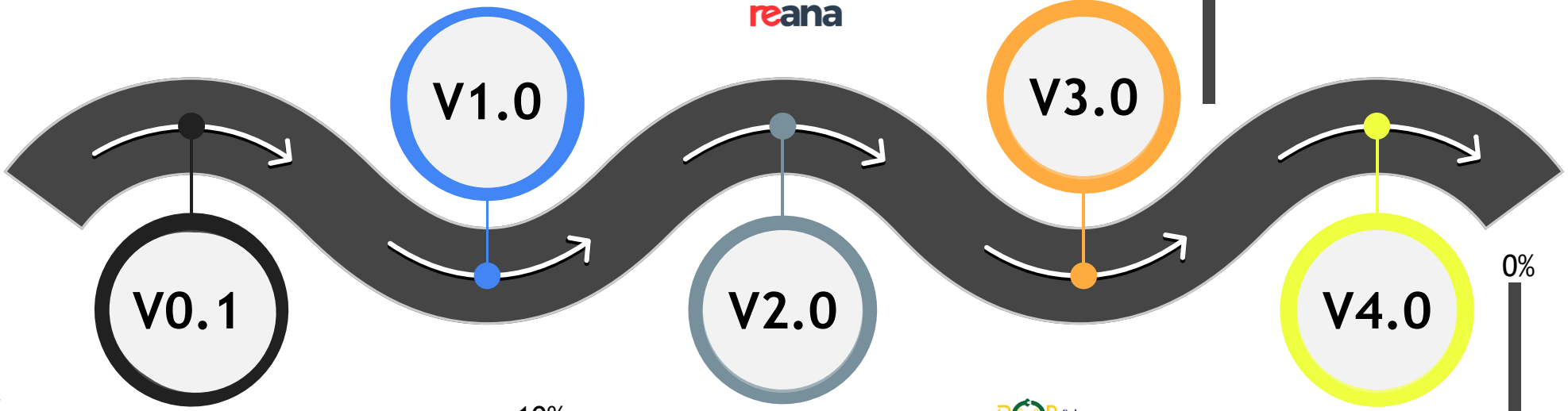
How does it look like?

Important links

https://aiod.eu/

- Form for service deployment on the AIoD - http://surl.li/pprlx
- Design assistance - https://aiod.eu/design-assistance
- Feedback for bug reporting on the AIoD - http://surl.li/pprpf
- Feedback on Release 1.0: http://surl.li/pprug

Development cycles based on releases



100%

Maintain current systems

Increase on Assets on Drupal and Eclipse Graphene

Improvements on the current navigation and menus

HUMANE AI NET

100%

Metadata catalogue
Open API - functionalities (GET, POST, DELETE)
1st Version of SDK
Deployments on kubernetes cluster
Migration plan of the AI Assets - Drupal, Eclipse Graphene
First set of connectors: Zenodo, Hugging Face, OpenML,
AI Builder - Cognitive architectures
First set of external services - not AI4Europe
Federated authentication
(MyLibrary) Marketplace
Services to substitute current system
- Additional experimentation services
- Reproducibility environment
- ICT 49 integration

OpenML

AI4 | eosc

reana

10%

Connectors to additional platforms
Distributed environment
Local and federated environment
Modules for inference and training
Extension of the AIoD connectors
Mechanisms for TAI support
Aggregation of dynamic HW resources to the resource pool.
Success Stories service

euROBIN

AIDA

TAILOR

STARLINGX

TBD

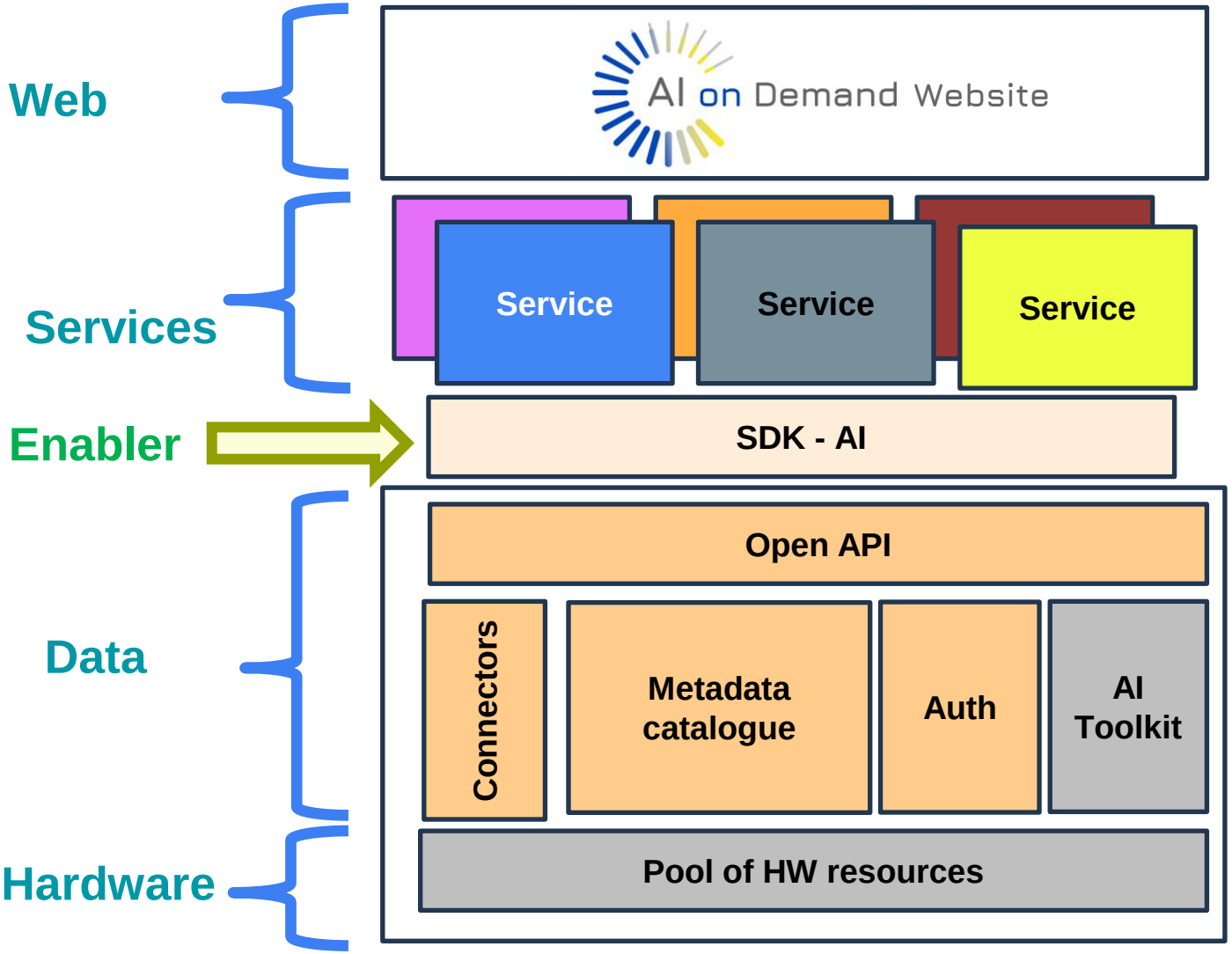
0%

TBD

0%

Funded by the European Union

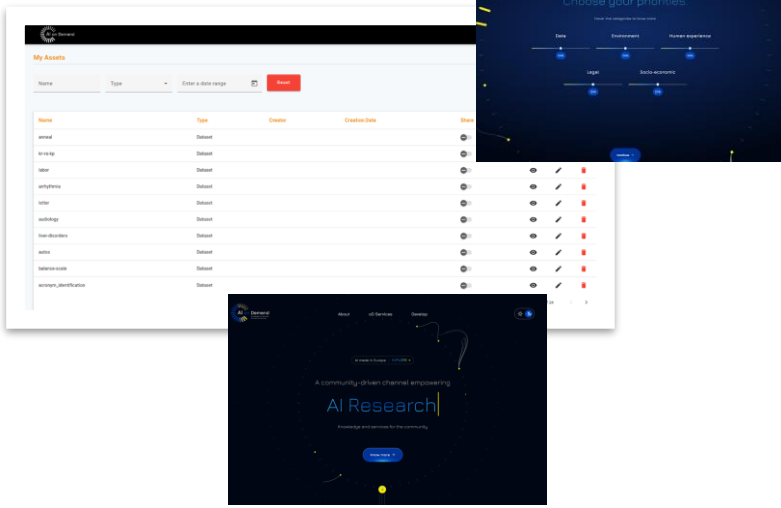
Sharing – Services, Data, Hardware

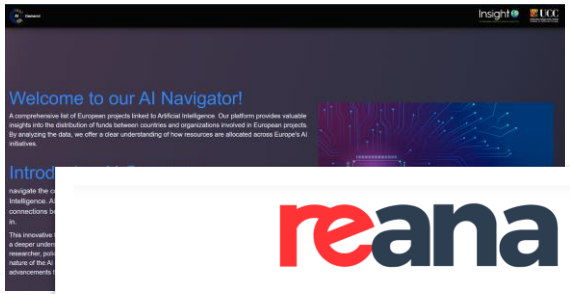


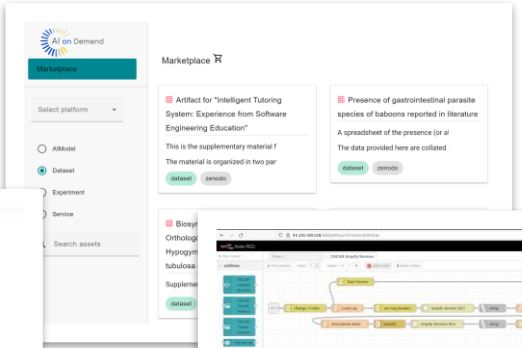
Funded by the European Union

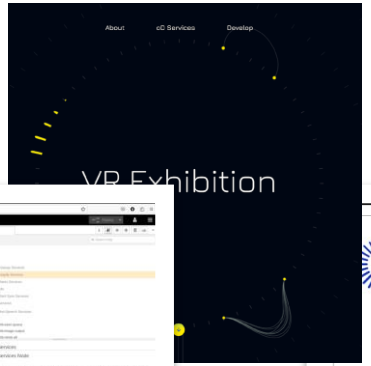
What we did

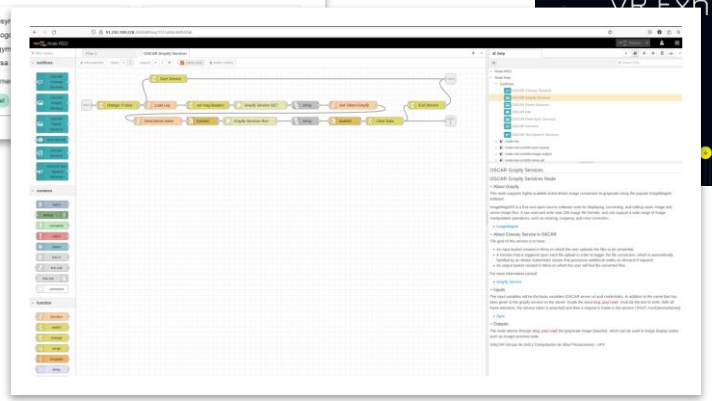
- Learn from the current systems and improve
- Make AIoD a system that provides the latest cutting edge AI Assets
- Integrate mechanisms to easily incorporate services (Assets and SaaS) into the platform

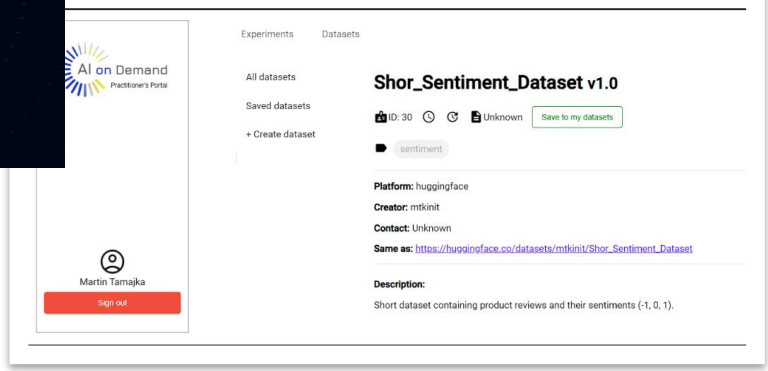








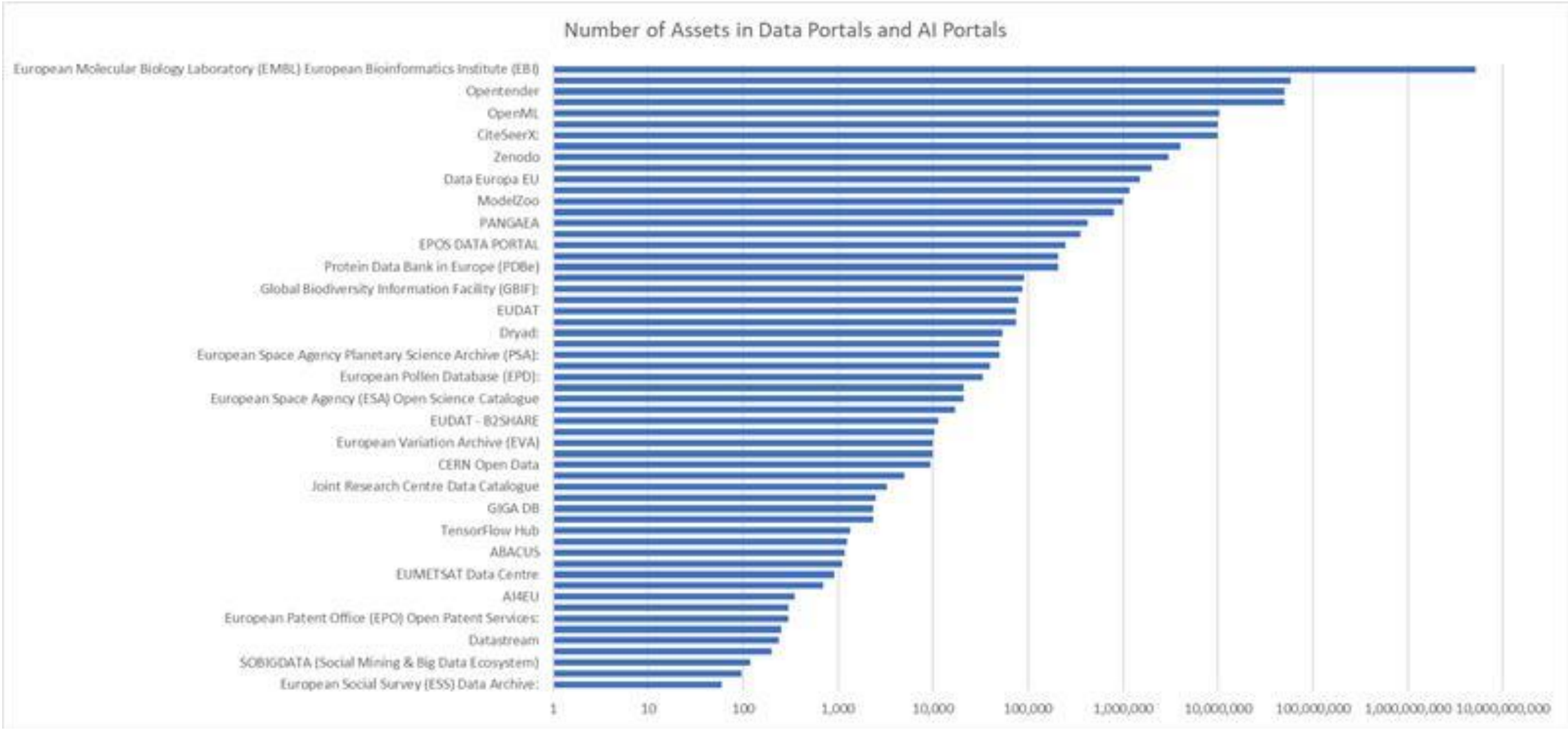




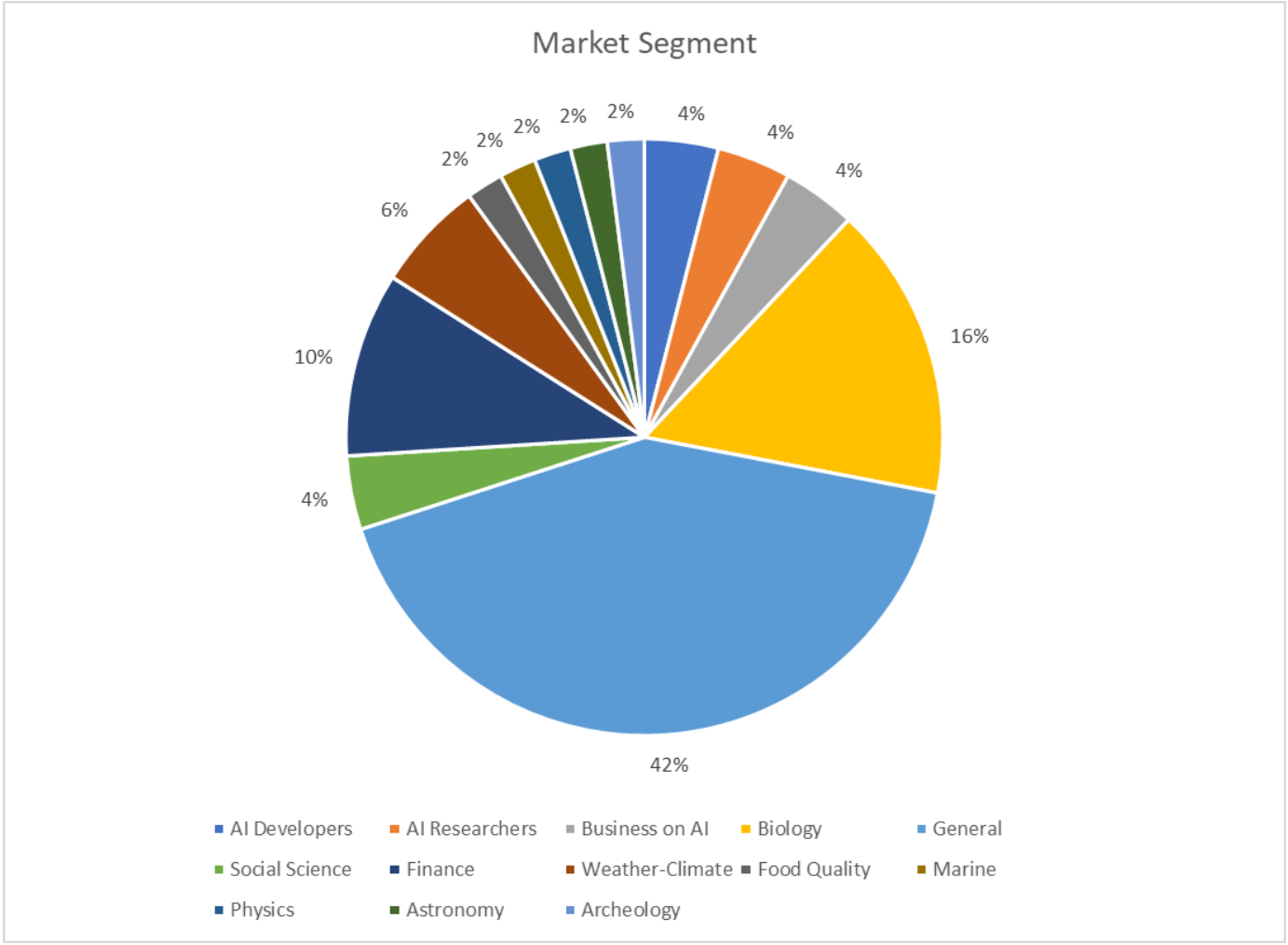
Let’s recap!

AI4EUROPE

Some of the EU platforms with AI* Assets



The market segments of those AI* Assets

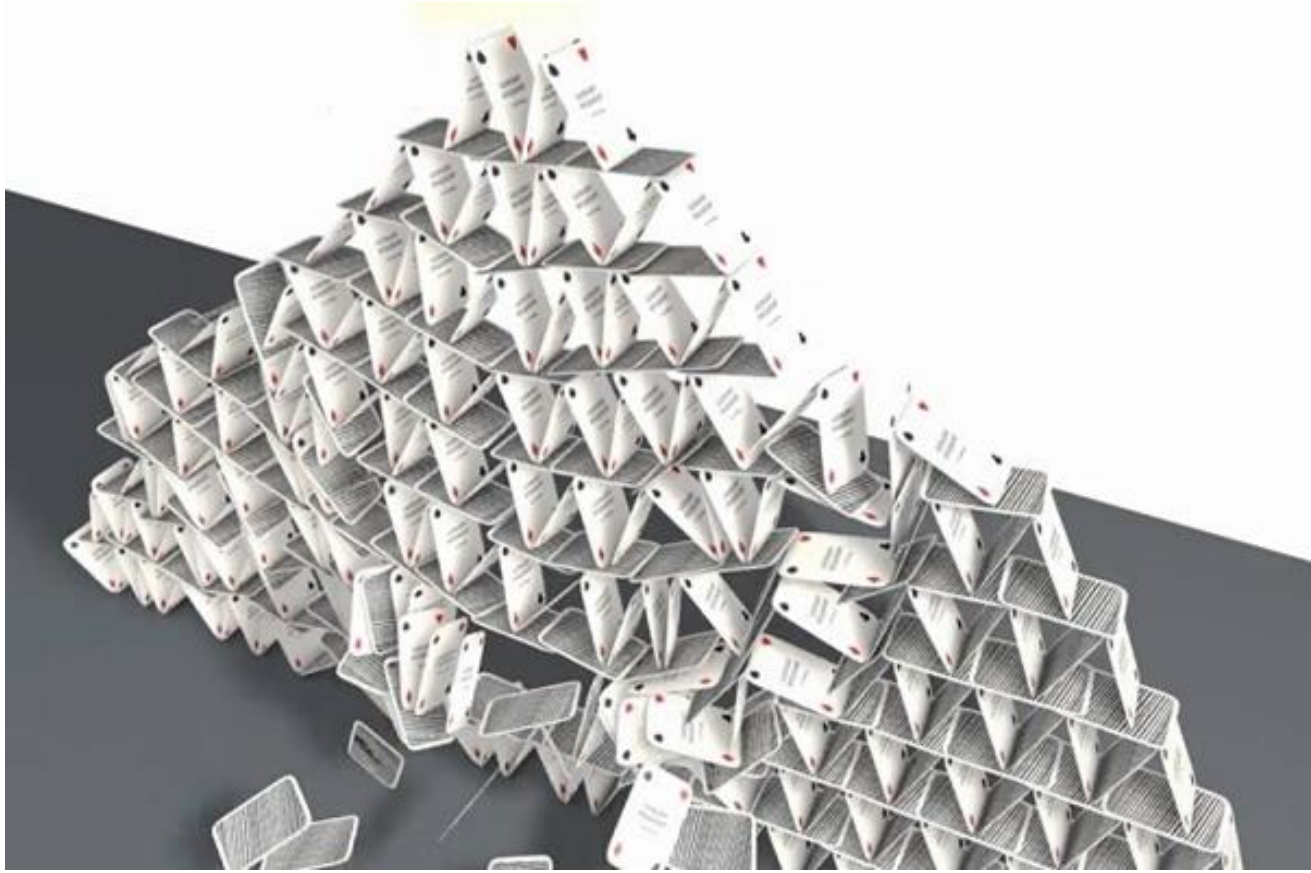


Analysis of the systems
(What did we realise?)

AI4EUROPE

Lessons learnt

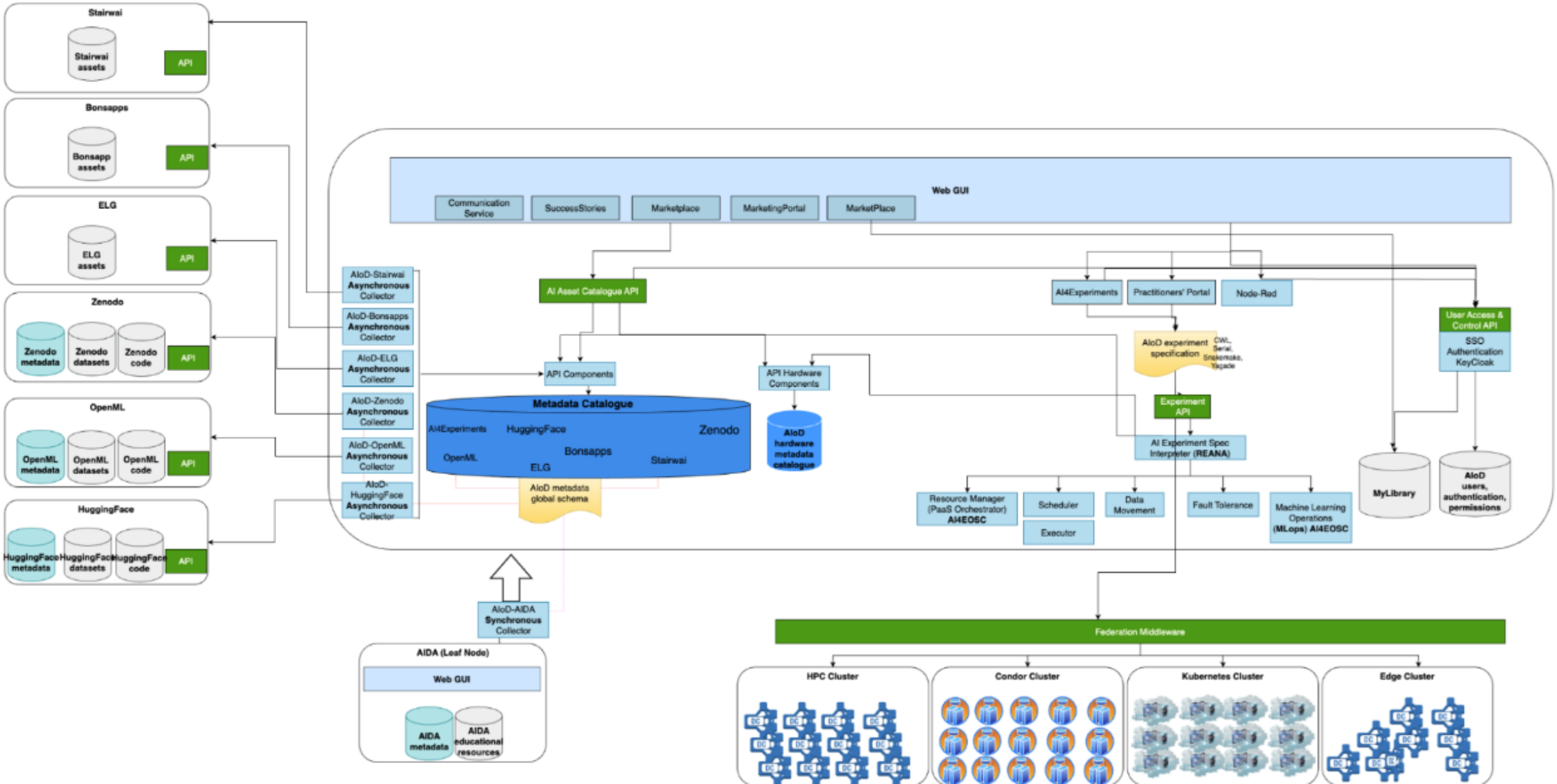
- Systems must scale
- Easy to incorporate new services (SaaS) that are novel and do not fit into any of the two subsystems (Drupal, Eclipse Graphene)
- Allow developers to use their preferences with coding languages
- To preserve identity of contributors (projects, associations, NoEs, ..)



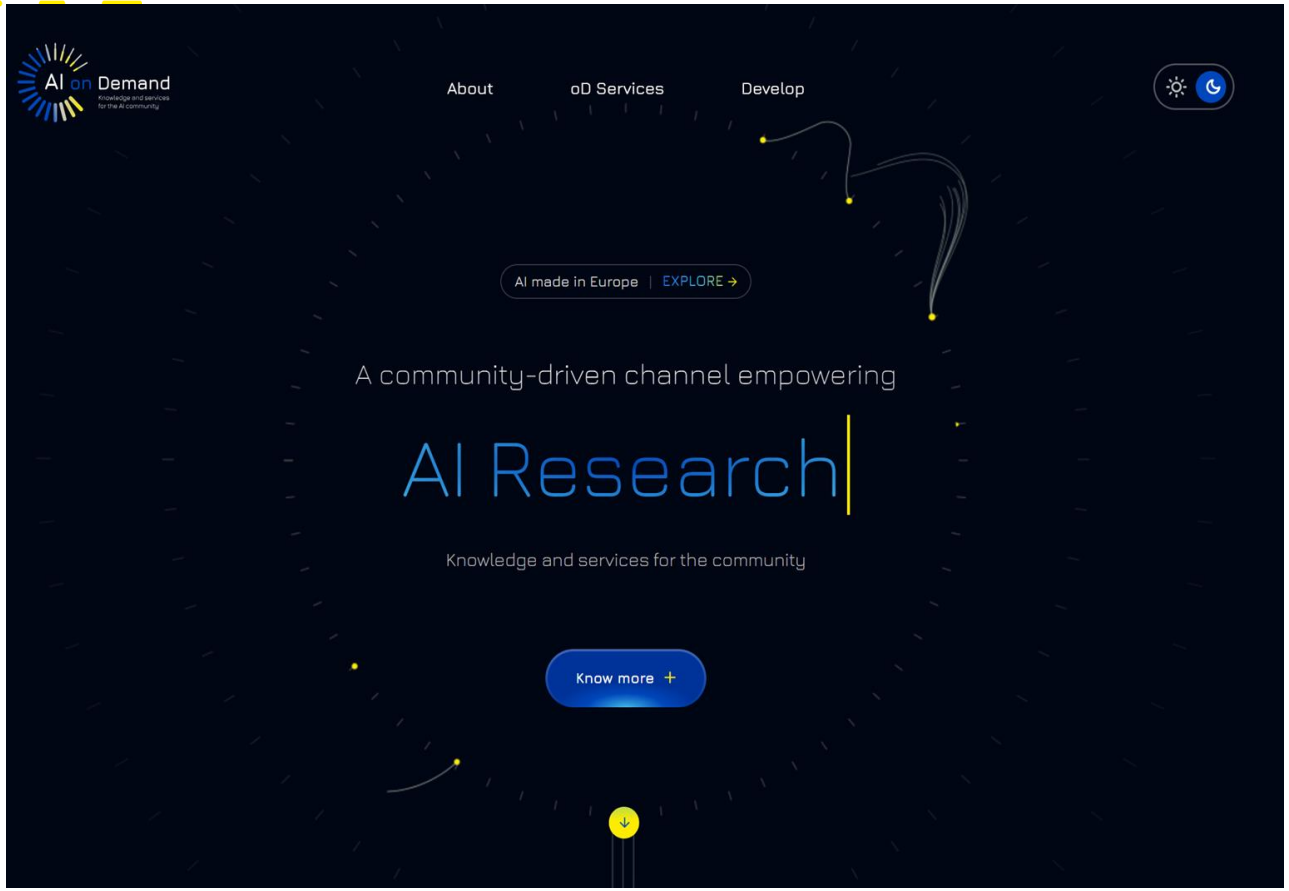
Can we do it better?

AI4EUROPE

The AIoD Architecture



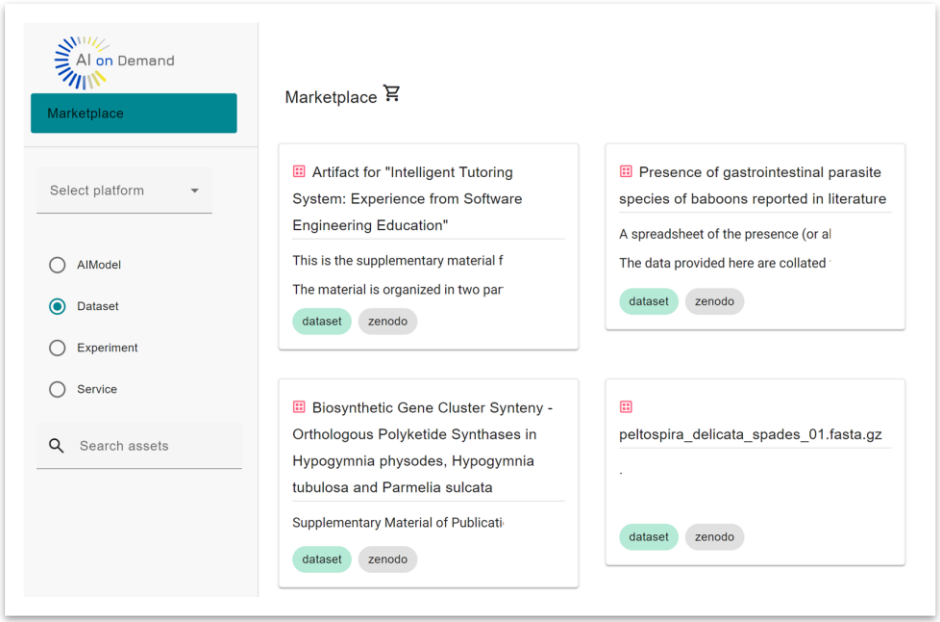
The Portal

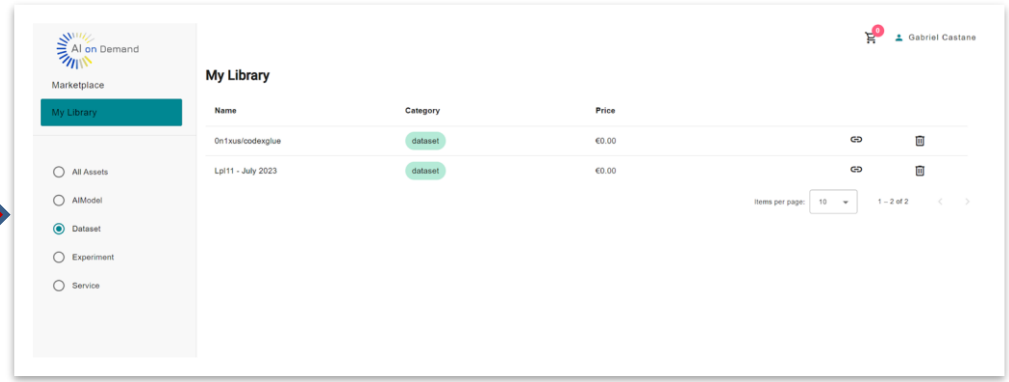


- System detached from platform
- Focused on marketing and promotion of systems
- No users or catalogues management
- Simplified navigation service-based
- Two backgrounds dark/light

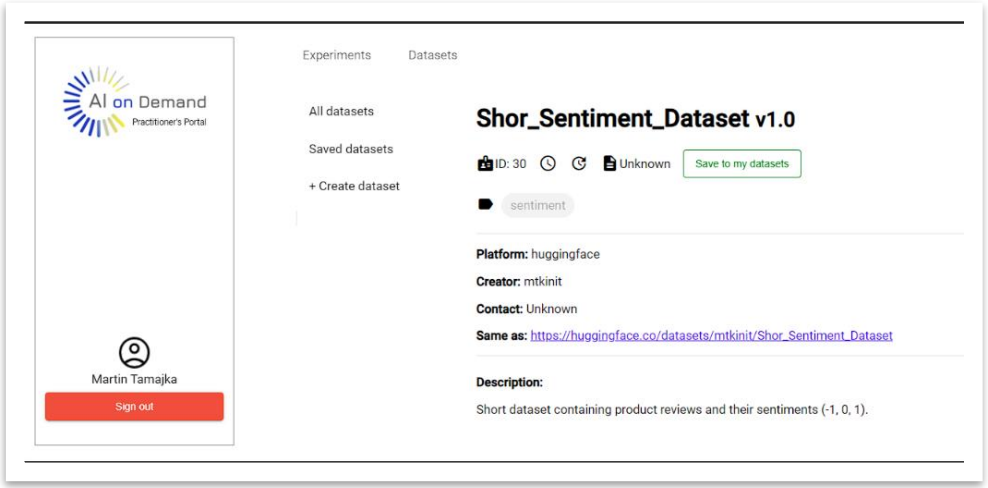
My Library

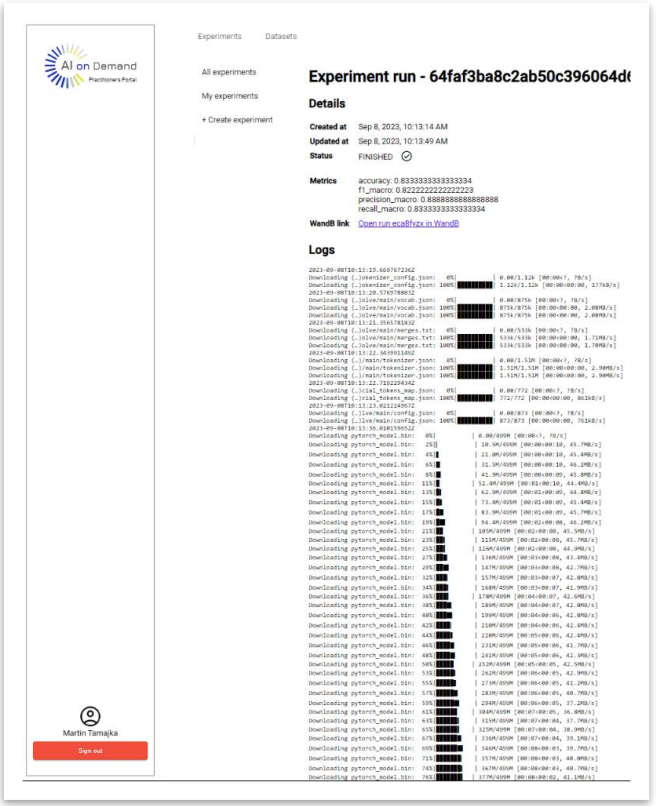
- Mechanism to access the AI Assets on the platform from a UI.
- User can create their own collection of Assets and Services.





RAIL (Example of a service)



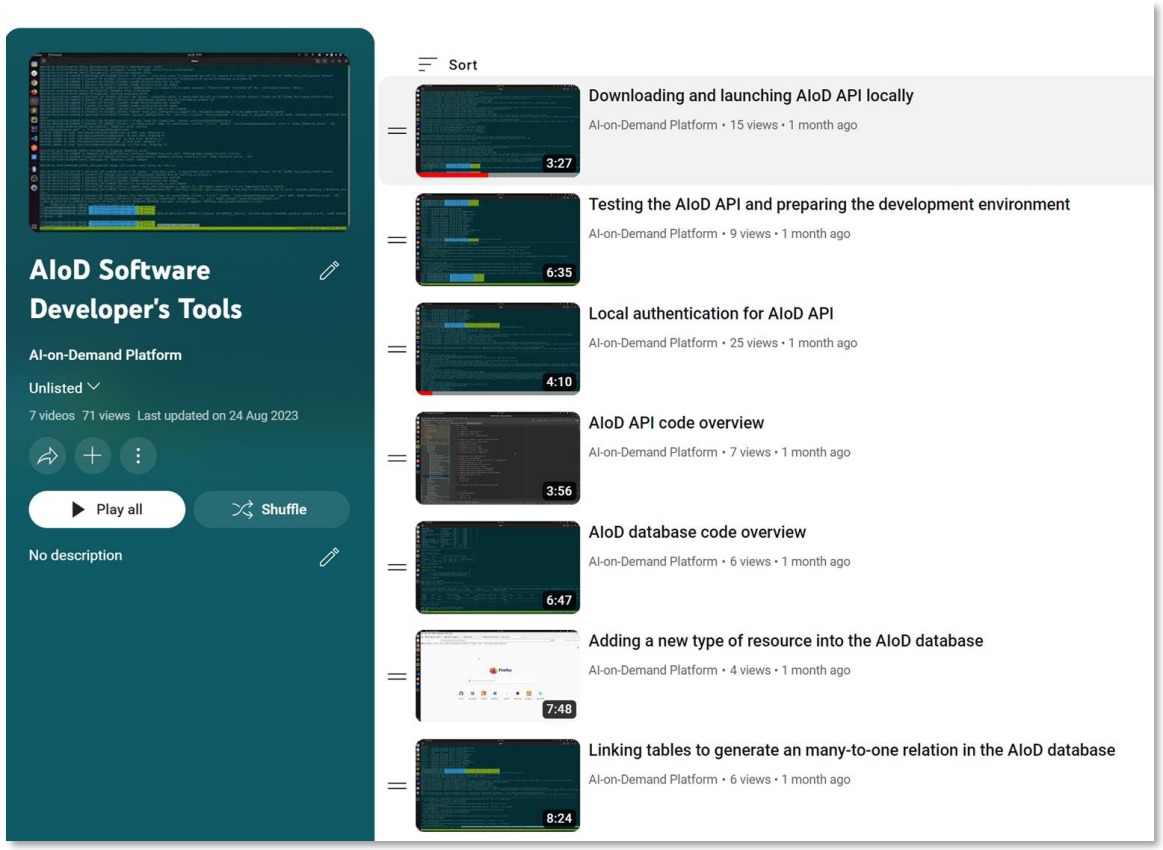


- Experimentation environment to create tests and executions from the metadata catalogue
- Combinations of experiments and datasets, and exploring results
- Potential for connection with other projects – Fairness, planning, etc.

Open API and authentication service

- The open API and metadata
  - Swagger - https://api.aiod.eu/
  - The authentication mechanism
    1. https://test.openml.org/demo uses EGI Check-in
    2. https://ai4europe.test.fedcloud.eu/ uses Keycloak federated with EGI Check-in
    3. https://auth.aiod.eu/aiod-auth/ uses keycloak federated with EGI Check-in

For developers



- Collection of videos on AIoD youtube channel https://www.youtube.com/playlist?list=PLL80pOdPsmF6hXGJVKT2v_-QUv-FFnwmc
- 7 videos – to continue growing and updating.
- To create videos for the use of services

Code SDK and platform

- Github - https://github.com/aiondemand/

Analysis of the idea - competitors

	OpenML	Eclipse Graphene	HuggingFace	ChameleonCloud's Trovi	NAIRR	Code Ocean	Azure ML (and other providers)
Purpose	MachineLearning	Machine Learning and hybrid pipelines	Machine Learning	Any algorithm that can be implemented in Jupyter notebook	Any AI Algorithm	In silico experiments	Machine Learning
Access to Datasets	5.5k	120	57.6K	None	Aggregates data from many different sources: data.gov (250k), NASA open data (10k), noaa big data (+100), nih data repositories (+200), nist science data (1k), patent and trade datasets (13.2M)	Datasets cannot be accessed directly, but through the code.	27
Access to AI assets	16.7K AI models & pipelines	173	306.2K AI models	96 artifacts but not all related to AI	Unknown, under development	It is focused on any in-silico experiment, including physics, computer science, bioinformatics, etc. Some of them	Foundation AI models + Aggregates HuggingFace
Technology for the design of experiments	Python	Graphical programming language	Python	Python Jupyter notebook	Under development, unknown	Any programming language and graphical programming language	Graphical programming language: Azure Machine Learning Designer
Execution environments	Users' computing resources	Small Fraunhofer cluster and users' computer resources	AWS / Azure	NSF funded computing resources	NSF funded computing resources, including at least 1 supercomputer	CodeOcean and AWS	Microsoft Azure
Discoverability	REST API / Web	REST API / Web	REST API / Web	Web	Under development, unknown	Web	REST API / Web
Licencing	Open source	multilicence	multilicence	multilicence	Under development, unknown	no	multilicence

First steps - Resources

	OpenML	Eclipse Graphene	HuggingFace	ChameleonCloud's Trovi	NAIRR	Code Ocean	Azure ML (and other providers)	AIoD
Purpose	MachineLearning	Machine Learning and hybrid pipelines	Machine Learning	Any algorithm that can be implemented in Jupyter notebook	Any AI Algorithm	In silico experiments	Machine Learning	**Any AI algorithm + DevOps**
Access to Datasets	5.5k	120	57.6K	None	Aggregates data from many different sources: data.gov (250k), NASA open data (10k), noaa big data (+100), nih data repositories (+200), nist science data (1k), patent and trade datasets (13.2M)	Datasets cannot be accessed directly, but th	27	**Aggregates OpenML, Eclipse Graphene, HuggingFace, Zenodo, etc.**
Access to AI assets	16.7K AI models & pipelines	173	306.2K AI models	96 artifacts but not all related to AI	Unknown, under development	It is focused on any in-silico experiment, in computer science, bioinformatics, etc. Some of them	Foundation AI models + Aggregates HuggingFace	**Aggregates OpenML, Eclipse Graphene, HuggingFace, Zenodo, etc.**
Technology for the design of experiments	Python	Graphical programming language	Python	Python Jupyter notebook	Under development, unknown	Any programming language and graphical programming language	Graphical programming language: Azure Machine Learning Designer	**Through services and 3rd party (Practitioner's Portal supports any programming language, node-red supports a graphical programming language)**
Execution environments	Users' computing resources	Small Fraunhofer cluster and users' computer resources	AWS / Azure	NSF funded computing resources	NSF funded computing resources, including at least 1 supercomputer	CodeOcean and AWS	Microsoft Azure	**Sharing, Through EGI clouds, potential creation of a marketplace for HW resources**
Discoverability	REST API / Web	REST API / Web	REST API / Web	Web	Under development, unknown	Web	REST API / Web	**REST API / Web**
Licencing	Open source	multilicence	multilicence	multilicence	Under development, unknown	no	multilicence	**Automatic licencing open source + commercial**

Analysis of the idea – SDK on AI (i)

SDK	Common Use	Programming Language(s)	Strengths	Weaknesses
AIMMS	Optimization modeling and decision analytics	AIMMS Modeling Language	Focuses on optimization and decision analytics, versatile for modeling, user-friendly IDE	Less suited for deep learning and neural network-based AI tasks, primarily focused on optimization and analytics
AllenNLP	Natural language processing, research	Python	Specialized for NLP, modular and extensible, strong for academic research, active development and community support	May not offer as broad a range of AI tasks as general-purpose frameworks, more research-oriented
AllenAI	Natural language processing, research	Python	Specialized for NLP, powerful for research, integration with various NLP tasks, active development and community	May not provide as much pre-built functionality as some commercial platforms, primarily research-oriented
MALLET	Machine learning for text analysis	Java	Open-source, versatile for text analysis, strong for topic modeling, comprehensive suite of tools	Command-line interface, may have a steeper learning curve for some users, fewer graphical interfaces
Gluon	Deep learning, versatile interface	Python, other languages	Flexibility, ease of use, interface for building and training models, supported by Apache MXNet and Microsoft CNTK	Smaller ecosystem compared to TensorFlow and PyTorch, primarily serves as an interface for other frameworks
MXNet	Deep learning, neural networks	Python, C++, Julia, R, Scala, Perl	Flexibility, multi-language support, scalability, efficient execution on multiple platforms	Smaller community compared to TensorFlow and PyTorch, less extensive pre-trained model availability
Apache OpenNLP	Natural language processing (NLP)	Java	Open-source, wide range of NLP tools, extensible, supports multiple languages, part of the Apache ecosystem	May have a steeper learning curve for Java developers, less extensive pre-trained models
Apache Mahout	scalable machine learning frameworks		It aids in clustering, collaborative filtering, and classification. Its computational operations make use of Java libraries, which are faster.	Python libraries are not as compatible as Java libraries with this framework. Its computational operations are slower than Spark MLib.
Spark MLlib	Distributed machine learning	Multiple (Scala, Java, Python)	Integration with Apache Spark, scalability, distributed data processing, comprehensive set of machine learning algorithms	Focuses on distributed machine learning, may not have the same breadth of deep learning capabilities as specialized frameworks
Core ML (for iOS)	Mobile app integration of ML models	Swift, Objective-C	Integration with iOS and macOS, simplified deployment of ML models in Apple apps	Limited to Apple's ecosystem, may not be suitable for non-iOS projects
Turi Create	Machine learning in Python	Python	Simplifies model development, supported by Apple, productivity-focused, user-friendly	Limited to Python, may not offer the same depth as specialized deep learning frameworks
PaddlePaddle (Paddle)	Deep learning, NLP, computer vision	Python, C++, and more	Versatile platform, support for various AI tasks, open-source, strong presence in China, easy-to-use API	Smaller global community compared to TensorFlow and PyTorch, less extensive pre-trained model availability
Caffe / Caffe2	Computer vision, deep learning	C++, Python	Speed and efficiency for CNNs, popular in computer vision tasks, pre-trained models available	Less flexibility for custom network architectures, steeper learning curve for some users
Clarifai	Computer vision, image and video analysis	Python, Java, Node.js	AI-powered image and video analysis, customizable models, cloud-based API, ready-made solutions for various domains	May involve cloud service costs, less focus on non-vision AI tasks
XGBoost	Gradient boosting for machine learning	Multiple (Python, R, Java)	High-performance gradient boosting library, excellent for tabular data, wide language support	Focuses on gradient boosting, may not be as suitable for deep learning or neural networks
Deeplearning4j	Deep learning, Java applications	Java, Scala	Java-focused, compatibility with JVM, scalability, integration with Hadoop	May have a steeper learning curve for Java developers, smaller community compared to Python-based frameworks
PyTorch	Deep learning, dynamic computation graph	Python	Dynamic computation graph, strong research adoption, user-friendly, growing community	Historically, limited production deployment support, smaller ecosystem compared to TensorFlow
ROBERTA (A Robustly Optimized BERT Pretraining Approach)	Natural language processing, text understanding	Python	State-of-the-art NLP model, strong performance on a range of NLP tasks, easy integration with Hugging Face Transformers	Specialized for NLP tasks, may not provide solutions for non-language-related AI tasks

Analysis of the idea – SDK on AI (ii)

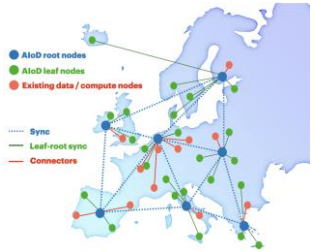
SDK	Common Use	Programming Language(s)	Strengths	Weaknesses
PyText	Natural language processing, text classification	Python	Specialized for NLP, deep integration with PyTorch, efficient for text classification, designed for production use	May not offer the same breadth of NLP tasks as some other frameworks, primarily for text-related tasks
FastAI	Deep learning, education, NLP	Python	Educational focus, easy-to-use, top-notch documentation, provides high-level abstractions for deep learning	May not be as versatile for production-level deployment, primarily suited for educational purposes
Keras	Rapid prototyping of neural networks	Python	High-level API, user-friendly, great for beginners, modular and extensible	Less flexibility for custom models, may require switching to TensorFlow or PyTorch for advanced use cases
TensorFlow	Deep learning, neural networks	Python, C++, Go, Java, Swift	Scalability, extensive community support, deployment options, hardware support (TPU), high flexibility	Steeper learning curve, verbosity, sometimes complex API
Dialogflow	Natural language understanding, chatbots	Multiple (Node.js, Java, Python)	Cloud-based, integration with Google Cloud, multi-language support, ready-made chatbot capabilities	Limited customization, may involve cloud service costs, primarily suited for chatbot applications
Google ML Kit	A mobile SDK that brings Google's machine learning capabilities to Android and iOS apps, with powerful, yet easy-to-use solutions	Java (Android), Swift/Objective-C (iOS)	Easy to integrate into mobile apps; Pre-trained models available; Supports custom TensorFlow Lite models	Limited to mobile app development; Some advanced use cases may need further fine-tuning.
Hugging Face Transformers	Natural language processing, NLP, chatbots	Python	Extensive collection of pre-trained language models, user-friendly, active community, great for NLP research	May not be as feature-rich for non-NLP tasks, focused on NLP and language understanding
IBM Watson Studio	Data science and AI development platform	Python, R, Scala, and more	Integrated environment, tools for data preparation, model development, and deployment	Proprietary, may involve licensing costs, primarily tailored to the IBM ecosystem
IBM Watson Machine Learning	Machine learning and AI on the cloud	Python, R, Scala, and more	Cloud-based, auto-scaling, model monitoring and management, integration with various data sources	Proprietary, may involve cloud service costs, primarily tailored to the IBM ecosystem
Watson Assistant	Conversational AI and chatbot development	Multiple (Node.js, Java, Python)	Cloud-based, integration with Watson services, multi-language support, ready-made chatbot capabilities	Limited customization, may involve cloud service costs, primarily suited for chatbot applications
Kaldi	Automatic speech recognition (ASR)	C++, Python	Open-source, specialized for ASR, powerful ASR capabilities, well-suited for speech and audio processing	May have a steeper learning curve, focuses on ASR, less versatile for general AI tasks
CNTK (Microsoft Cognitive Toolkit)	Deep learning, neural networks	C++, Python	Efficient deep learning library, support for Microsoft products, scalability, strong GPU support	Smaller community compared to TensorFlow and PyTorch, may not have as extensive third-party model support
Lobe	Custom computer vision models	Windows, macOS	User-friendly GUI for model creation, integration with Lobe app, no coding required, quick model prototyping	Focuses primarily on computer vision, less control for advanced AI development
ML.NET	Machine learning in .NET applications	C#, F#, VB.NET	Integration with .NET ecosystem, support for various data sources, user-friendly API, offline model execution	Less extensive third-party library support, may require additional components for deep learning
MindsDB	Automated machine learning	Python, REST API	Simplified machine learning, focuses on AutoML, user-friendly, accessible through a REST API	Limited to automated machine learning, may not offer as much control as traditional ML frameworks
Theano	Deep learning, neural networks	Python	Efficient mathematical expression evaluation, strong GPU support	Less active development and community compared to TensorFlow and PyTorch, lower-level compared to high-level APIs. it's no longer actively maintained or updated.
Mycroft	Open-source voice assistant platform	Python, JavaScript	Open-source, extensible, privacy-focused, voice assistant capabilities, community-driven development	May require more development effort for extensive customization, less pre-built functionality than major voice assistants
NVIDIA Deep Learning SDK	Deep learning, GPU optimization	C++, CUDA, Python	GPU acceleration for deep learning, optimized for NVIDIA GPUs, high-performance, comprehensive GPU libraries	Tied to NVIDIA hardware, may not be as suitable for projects without NVIDIA GPU resources

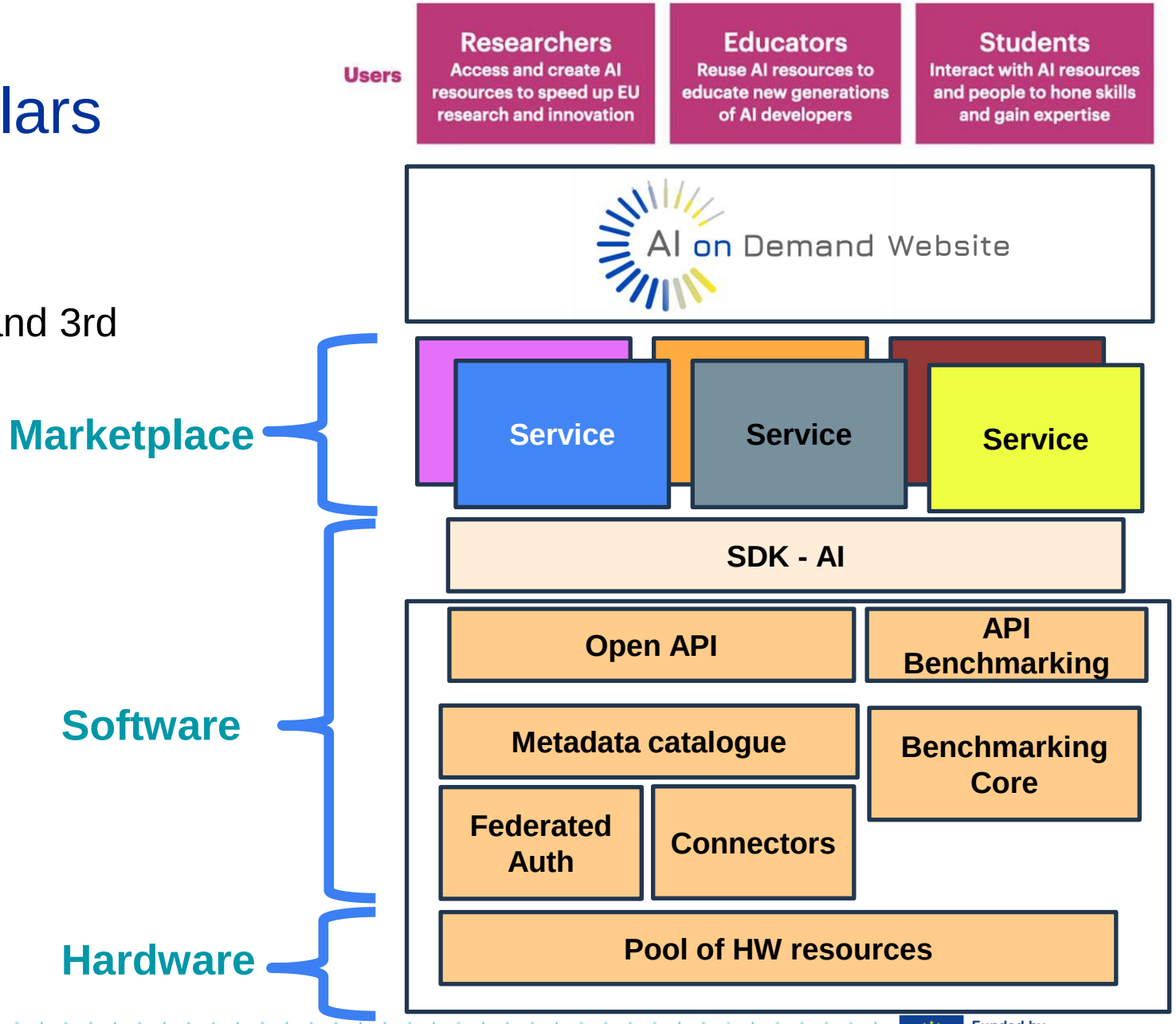
Analysis of the idea – SDK on AI (iii)

SDK	Common Use	Programming Language(s)	Strengths	Weaknesses
OpenAI GPT (Generative Pre-trained Transformer)	Natural language processing, text generation	Multiple (Python, JavaScript)	Cutting-edge NLP models, text generation capabilities, well-suited for creative and language-focused applications	Access may be limited, resource-intensive for large-scale use, primarily suited for language-related tasks
Scikit-learn	Traditional machine learning	Python	Easy to use, extensive library for machine learning algorithms, well-documented	Limited support for deep learning, primarily focused on traditional ML tasks
OpenCV	Computer vision, image processing	C++, Python, Java	Extensive library for computer vision tasks, wide range of image processing functions	Limited support for deep learning, more focused on computer vision and image processing
ONNX (Open Neural Network Exchange):	Model interoperability, compatibility	Multiple	Facilitates model exchange between different frameworks, ecosystem-agnostic	Not a deep learning framework itself, may require additional frameworks for model training and deployment
Plexaderm	Computer vision and image analysis	Python	Simplified computer vision, user-friendly interface, support for various computer vision tasks, quick model prototyping	Less control and versatility for advanced computer vision tasks, primarily for image analysis
Chainer	Deep learning, research, dynamic graphs	Python	Dynamic computation graph, flexibility, popular for research, easy-to-understand API	Smaller community compared to leading frameworks, may not be as suited for production-level deployment
Pycaret	Automated machine learning	Python	Simplified machine learning, supports various algorithms, automatic feature engineering, easy-to-use API	Limited to automated machine learning, may not provide the same level of control as manual model development
Gensim	Topic modeling and word vector representation	Python	Specialized for word embeddings and topic modeling, lightweight, efficient for natural language tasks	May not offer as much control for other AI domains, focuses primarily on text analysis
Rasa	Conversational AI, chatbots	Python	Open-source, specialized for conversational AI, flexible and customizable, strong NLP capabilities	May require more development effort for complex use cases, less emphasis on other AI domains
Snips	Offline voice recognition, privacy-focused	Python	Local voice recognition, privacy-centric, suitable for edge and IoT devices, customizable voice assistants	Limited language support, focus on specific voice recognition tasks, less extensive third-party integration
Acumos AI	Model sharing and deployment platform	Multiple	Simplifies model sharing and deployment, open-source, designed for use in AI model marketplaces	May require additional frameworks for model development, less focus on model training
Keras.js	Deep learning in JavaScript (browser)	JavaScript	Running deep learning models in web browsers, JavaScript-based, compatibility with Keras deep learning framework	Limited to web browser environments, less suitable for server-side AI tasks
Horovod	Distributed deep learning	TensorFlow, PyTorch, MXNet	Simplifies distributed training, scalable, efficient for large models and datasets	Focused on distributed training, may not provide standalone deep learning framework capabilities
AutoML Libraries (e.g., AutoKeras, H2O.ai)	Automated machine learning	Python	Automation of model selection and tuning, user-friendly, quick model development	Limited to specific automation tasks, may not offer the same level of control as manual model development
BERT (Bidirectional Encoder Representations from Transformers)	NLP, sentiment analysis, text classification	Python	Pre-trained NLP model, state-of-the-art for language understanding, applicable to various NLP tasks	Specialized for NLP tasks, may require additional components for broader AI applications
PyBrain	?	Python	Open-source ML library for Python. It provides a simple and flexible environment for experimenting with various machine learning algorithms and is perfect for researchers, educators, and developers looking for a lightweight Python-based framework for exploring machine learning concepts	limited documentation and a smaller community compared to mainstream libraries. It may also lack some advanced features found in other frameworks.
Amazon Machine Learning	data analysis, training of the model, and evaluation.		You don't need to write a lot of code with this framework. Instead, it lets you interact with the AI-powered framework via APIs. Commonly used by data scientists, developers, and ML researchers.	It lacks flexibility as the entire framework is abstracted, so if you'd like to choose a particular normalization or machine learning algorithm, you can't. It also lacks data visualization.
Torch	machine learning, computer vision, signal processing, parallel processing, image, video, audio	LuaJIT	Very high flexibility regarding languages and integrations High level of speed and GPU utilization efficiency Pre-existing models are available to train the data on.	Documentation is not very clear to the users, so it presents a steeper learning curve Lack of code for immediate use, so it takes time. It is initially based on a programming language called Lua, and not many are aware of it.

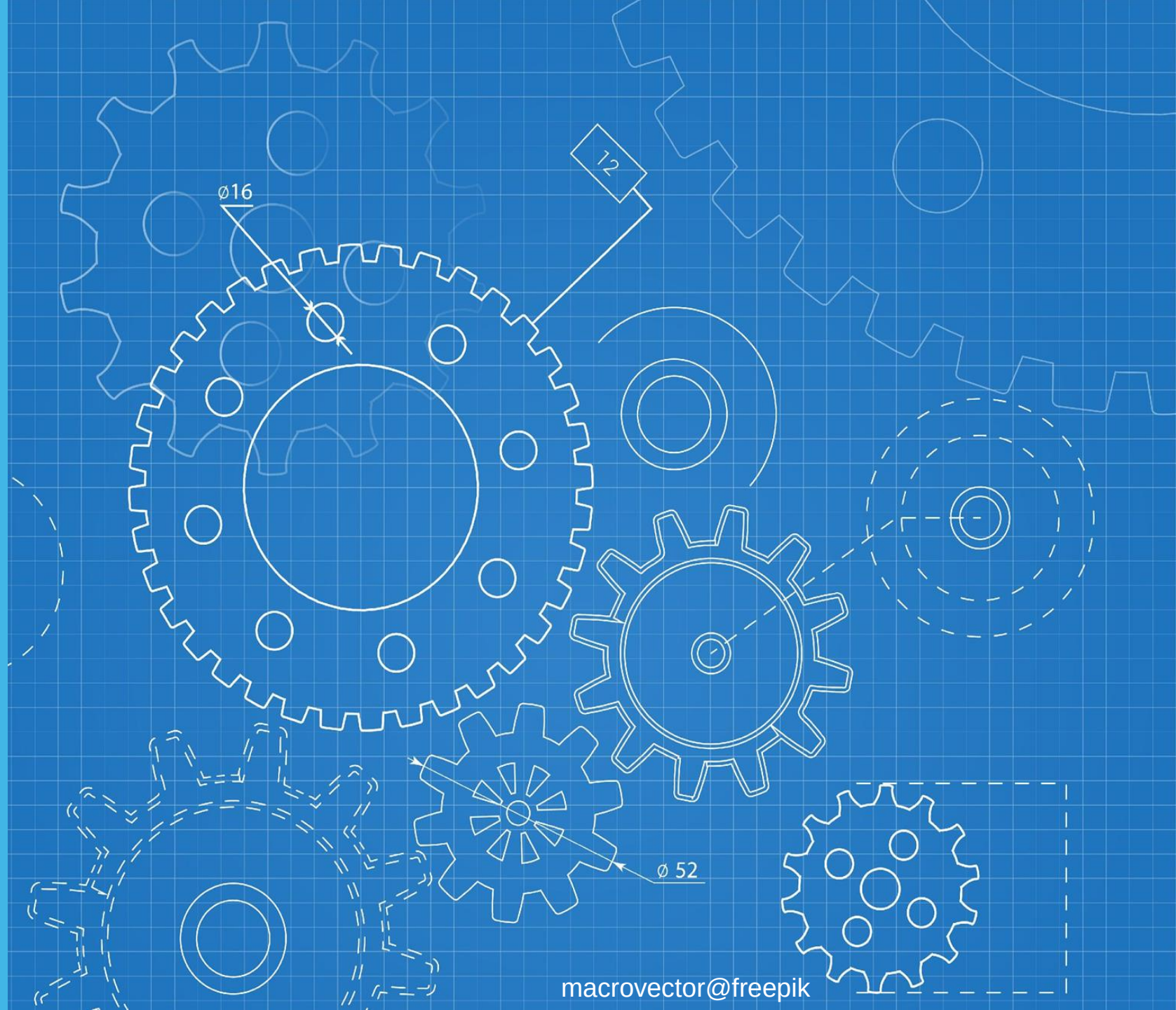
What are the main pillars

- A Portal
- Independent Services – AI4Europe and 3rd Parties
- For service integrators (developers)
  - Open API (Metadata catalogue)
  - Authentication





The process



macrovector@freepik

AI4EUROPE

New development cycles for AIoD based on releases

Maintenance

By service developers, and catalogues by ITI, TUE and UNIZAR.

Testing

Development on local environment, testing and preproduction on UNIZAR, EGI and TUE, ITI. ITI for production.

Collection of feedback on services and core architecture

Implementation

Agile methodology. Sprints starting by the core architecture, and followed by the services

Planning

Effort, migration of current system, risks, issues

Analysis

Analysis of the current platforms, benefits, advantages, gaps

Design

Group with 8 architects (different partners) designing the system

Working methodology

Communication

- Slack channels for development with AI4Europe and collaborators

Structure

9 Blocks

- Portal
- Migration
- Catalogue
- Authentication
- Synchronisation
- Administration
- Deployment
- Services
- Documentation

Knowhow

- Videos tutorials
- Documentation available
- Github repositories

DevOps environment

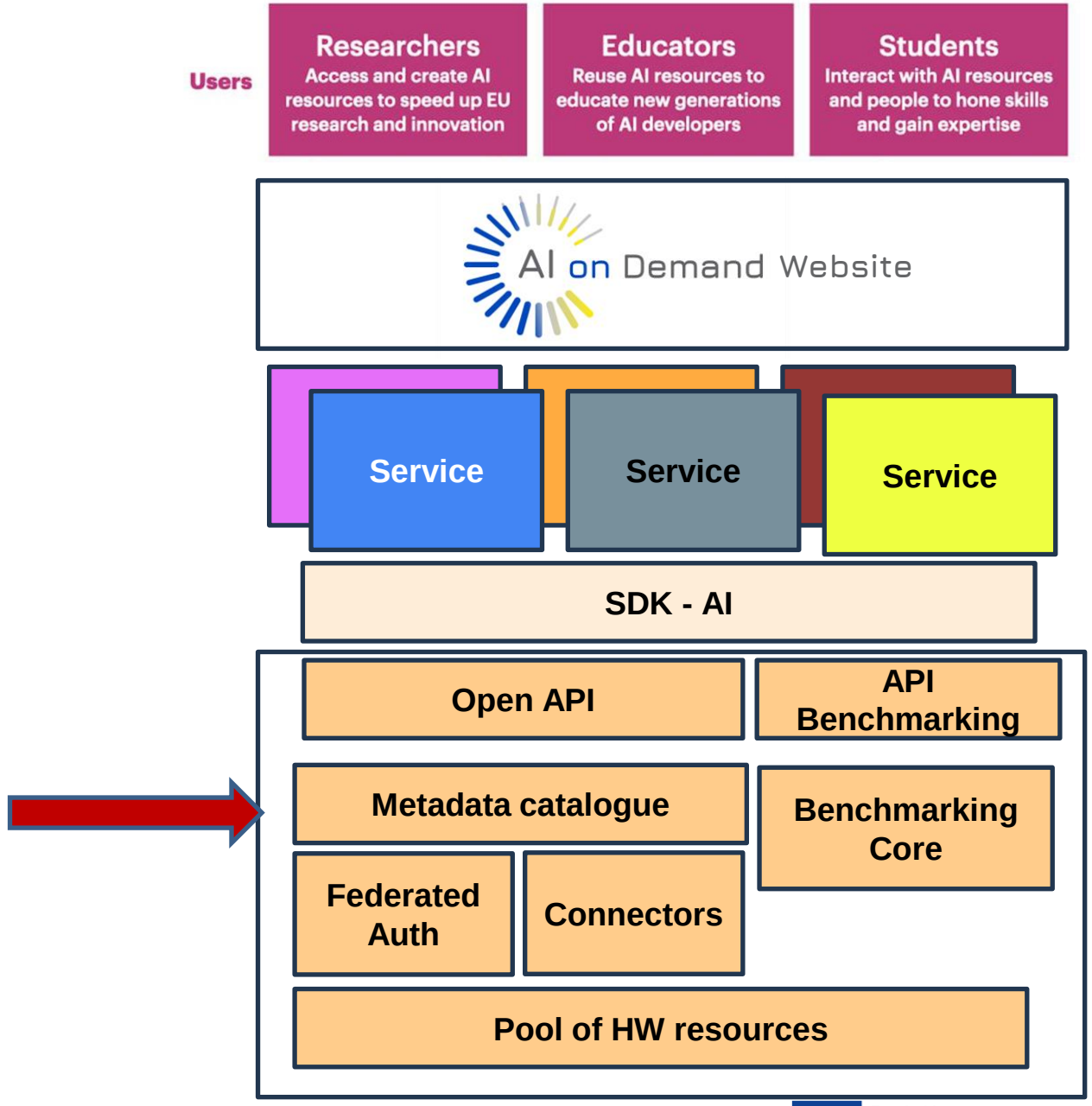
- Development, testing, preproduction and production
- 4 Nodes – ITI, TUE, EGI, and UNIZAR

Sprints

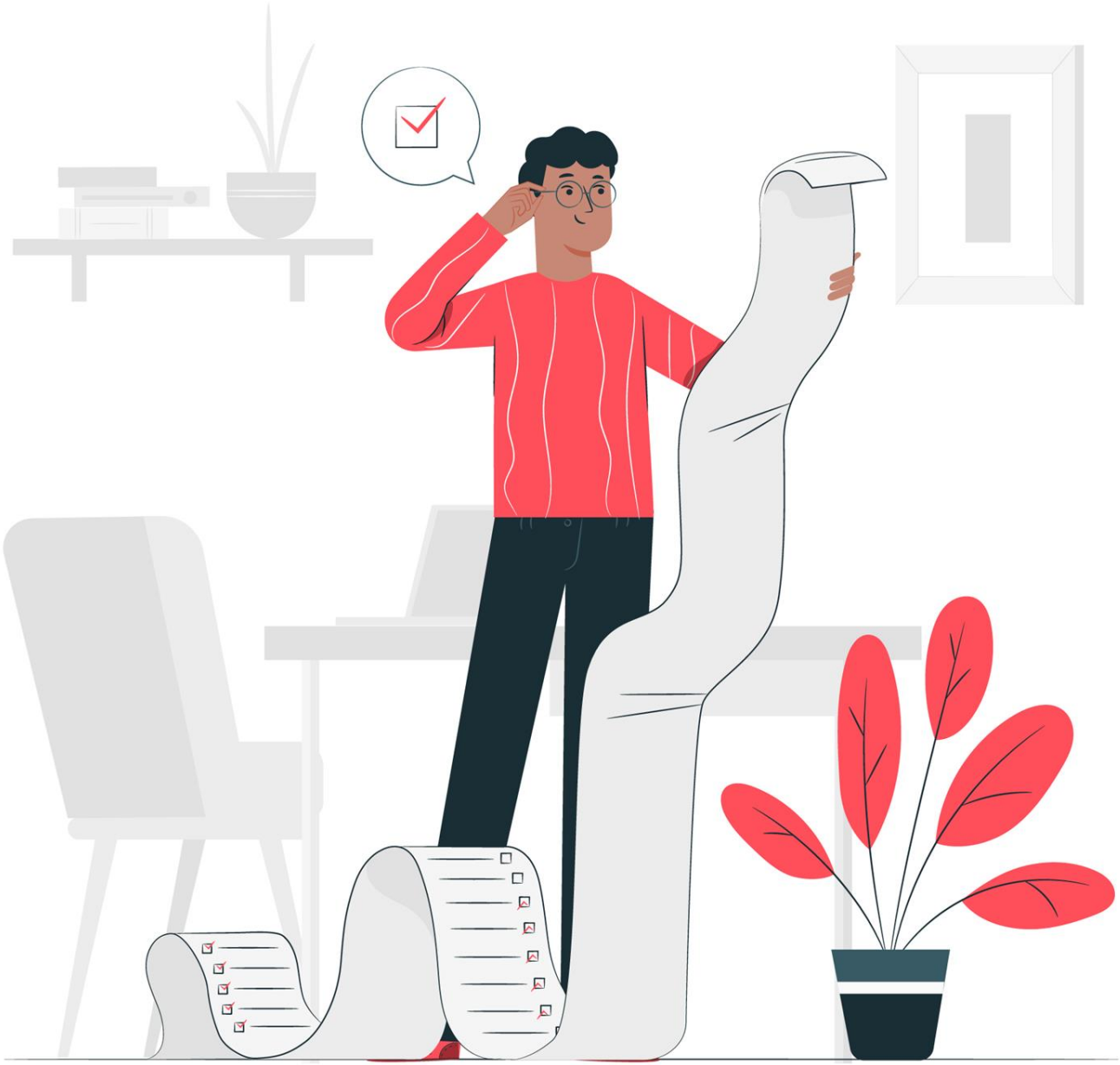
- 9 Sprints for all teams
- WP3, 4 and 5 cooperating together
- Technical coordination

Funded by the European Union

Migration

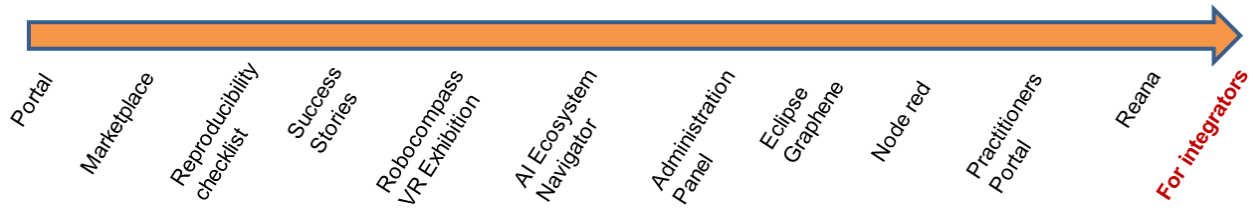

Users
Researchers
Access and create AI resources to speed up EU research and innovation
Educators
Reuse AI resources to educate new generations of AI developers
Students
Interact with AI resources and people to hone skills and gain expertise
AI on Demand Website
Service
Service
Service
SDK - AI
Open API
API Benchmarking
Metadata catalogue
Benchmarking Core
Federated Auth
Connectors
Pool of HW resources

Outline



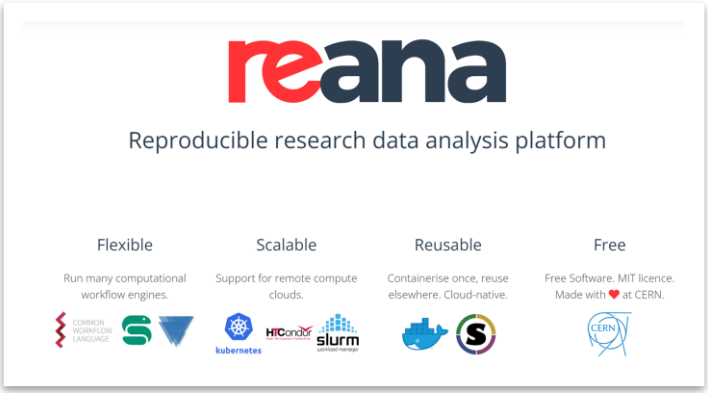
AI4EUROPE

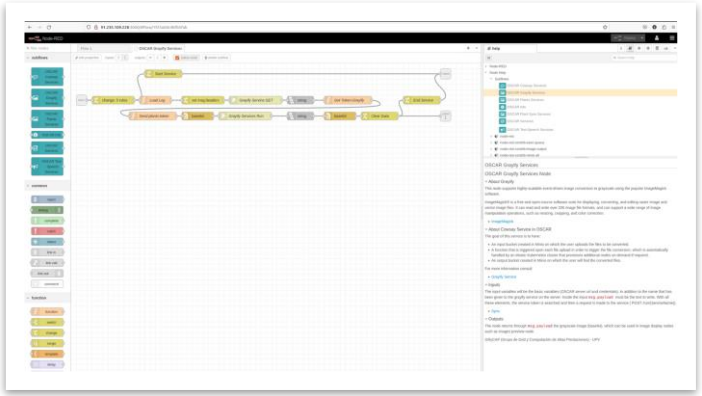
Brief summary

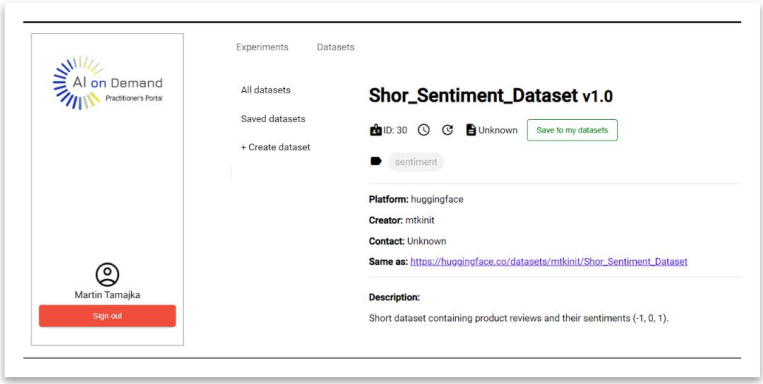


2 Operational systems inherited from AI4EU (391 assets and 329 exp)

- Open API and metadata catalogue with > 20 catalogues
- 4 testing and preproduction nodes
- 1 authentication mechanism (federated)
- 7 services contributed from external projects
- 3 services developed by AI4Europe
- 3 additional experimentation environments



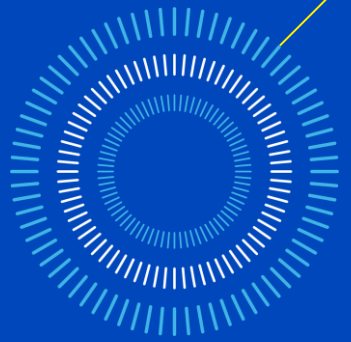




Second part: Hands-on session for practitioners

- **Metadata catalogue** - rest API (Jean Matias - UCC)
- **RAIL** (Martin Tamajka - KINIT)
- **AI Builder** (Sangamithra Panneer Selvam - FHG)

Thank you!



AI4EUROPE

Funded by
the European Union

Consortium


UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh


ÖREBRO UNIVERSITY


ITI
INVESTIGATE
TO INNOVATE


KINIT


telenor


ΔΗΜΟΚΡΙΤΟΣ
DEMOKRITOS


Universidad
Zaragoza
1542


Deutsches
Forschungszentrum
für Künstliche
Intelligenz GmbH


Fraunhofer
IAIS


eit
Digital


TU/e
EINDHOVEN
UNIVERSITY OF
TECHNOLOGY


ZRC SAZU


ARISTOTLE
UNIVERSITY
OF THESSALONIKI


ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA


BSC
Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación


Universiteit
Leiden
eLaw


egi


FGS


THE ARCTIC UNIVERSITY OF NORWAY
UIT


European
DIGITAL SME
Alliance


Université
Paris Cité


@
IeAT


LOBA®


A!
Aalto University

AI on Demand

Knowledge and services for the AI community

AI-on-Demand guidelines for communication and dissemination

AIoD Communication Board

Funded by the European Union

Index

Funded by
the European Union

“

The AI-on-Demand Platform (AIoD) is a community-driven channel designed to empower European research and innovation in Artificial Intelligence (AI), while ensuring the European seal of quality, trustworthiness and explainability”

Funded by
the European Union

Open and easily accessible, the AIoD facilitates knowledge sharing, research experimentation and development of state-of-the art solutions and technologies related with Artificial Intelligence.

The AIoD is for:

- ✔ AI researchers from academia or industry and students;
- ✔ SMEs and tech providers;
- ✔ Digital Innovation Hubs, EU funded projects and other EU bodies;
- ✔ Other AI and Technology enthusiasts.

Funded by
the European Union

The AIoD branding & how to use it

Funded by
the European Union

The AIoD branding

- The brand developed for the AI-on-Demand (AIoD) was conceived to embrace three main keywords: collaboration, community and growth.
- In a first layer, the lines in the symbol grouped in a circle represent cohesiveness, union, dynamism and expansion.
- In a second layer of the symbol, the gradient spiral represents development, interaction and circulation of knowledge, people and organisations.
- The colour grading reinforces the idea of development, progress and evolution. The shades of blue and yellow convey the “European feel”.

Funded by
the European Union

AIoD identity

- AI-on-Demand logo has three versions:

Logo with claim	Logo without claim	AIoD trademark
		

- The AIoD logo is available in different formats (.ai .jpg .png) for both printing or digital purposes.
- The brand manual available provides the specifications and guidelines when using the logo.

ACCESS HERE

Funded by the European Union

When to use the AIoD branding

- The AI-on-Demand identity can be used by any organisation or project to **communicate or promote** any activities, services, products or events associated with the AI-on-Demand (AIoD).
- The use of the AIoD trademark is optional but encouraged. It is up to the project or organisation to decide how to use it.
- Some examples for the integration are to:
  - Visibly place the AIoD trademark in the logo (check pg. 29 AIoD Brand Manual)
  - Include the AIoD trademark in footer of the website, banners, brochures/factsheets, publications etc.

Funded by
the European Union

AIoD Communication Toolkit

- AI-on-Demand **Communication Toolkit** can be used by the community to support any formal or informal communication activity where the AI-on-Demand should be represented.
  - Events/Meetings
  - Documents/Publications
- The Communication Toolkit is composed by:
  - Word template
  - Power point template
  - Folder
  - Letter head paper
  - Business cards

ACCESS HERE

Funded by
the European Union

The AIoD Promotional Material

- The AI-on-Demand **Promotional Material** can be used by the community to support the promotion of the AIoD contributing to its recognition and growth.
- AI-on-Demand promotional materials are composed by:
  - Brochure (digital + printing)
  - Flyer (digital + printing)
  - Badges
  - Stickers
  - Poster, roll-up and pop-up booth

ACCESS HERE

Funded by
the European Union

The AIoD outreach services

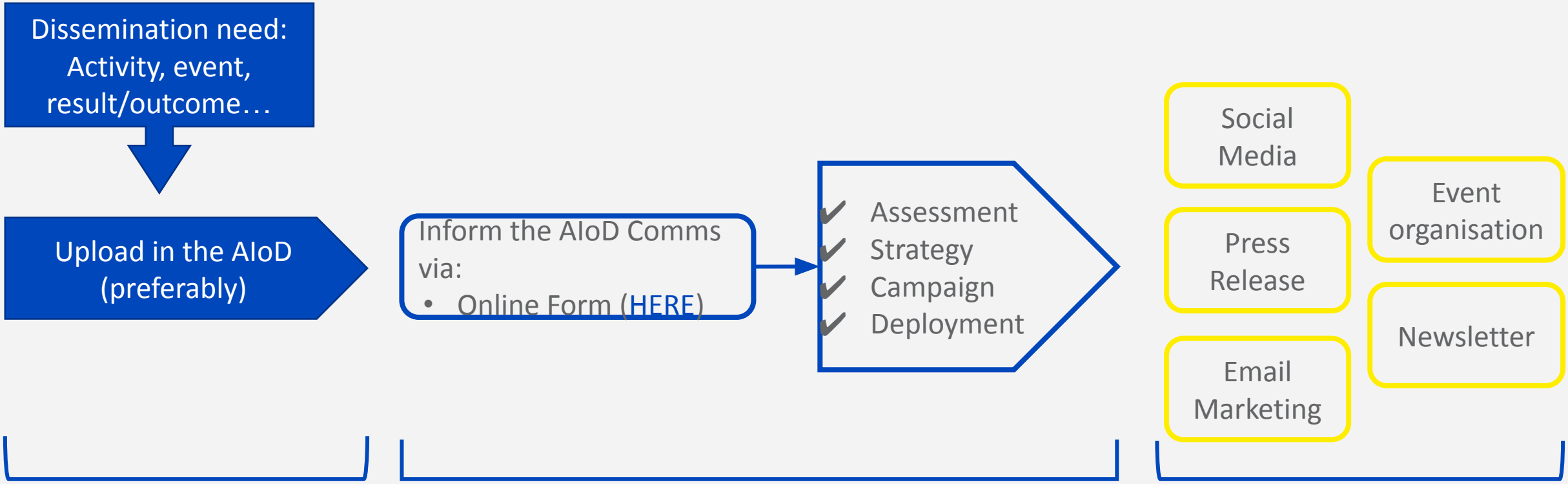
Funded by
the European Union

AIoD outreach services

- AIoD community (including European projects) can benefit from the visibility of AIoD channels.
- Support in the dissemination of activities, events, outcomes, etc can be requested to AIoD via an online form;
- Any information to be disseminated via AIoD channels should be first uploaded in the AIoD platform.

Funded by
the European Union

AIoD outreach services - roadmap



Funded by the European Union

Promotion in the AIoD Social Media @AIonDemand

- Activities, events and outcomes from the AIoD community can be promoted in the AIoD social media channels: X and LinkedIn, benefitting each one from a community of more than 4K followers.

- There are two ways for featuring their content:

  a) Mention AIoD (using the handle @AIonDemand) in the post published in project's social media channel. Then, AIoD communication board will validate the content and share it.

  b) Submit a request to the AIoD communication team via online form providing the information needed for creating the campaign.

- If there is the need to produce live content (for example, at events), some images and 1/2 sentences should be provided for the email comms@aiod.eu, so that we can communicate them as soon as possible.

Funded by the European Union

Featuring in the AIoD Newsletter & email marketing

- AIoD will release quarterly newsletters to its users and subscribers.
- Content from European funded projects can be featured in this newsletter.
- Content can be also distributed in specific email marketing campaigns.

Funded by
the European Union

Press release distribution

- AIoD can support the distribution of Press Releases through a database of journalists and media outlets.
- However, to ensure a successful media coverage it is recommended to establish close media relations, and this type of support the AIoD will not be able to provide.
- The press releases need to be developed by the projects.
- Media coverage is not guaranteed.

Funded by
the European Union

Support for event promotion or organisation

- European funded projects organising events will be able to request for support from the AIoD in the form of:

  a) Promotion: having the AIoD supporting as “promotional partner” - promoting the event across the various communication channels;

  b) Organisation: having the AIoD as co-organiser providing support in the:

    - Definition or setting up of the agenda/program.
    - Identification of speakers/experts, using the AIoD large community.
    - Organisation of joint events/sessions for the participation of several projects or clusters.

Funded by the European Union

Use of the AIoD platform for dissemination purposes

Funded by
the European Union

How to use and benefit from the AIoD platform for dissemination purposes

1. Create a **profile** of a project or organisation in the AIoD platform through the Contribution Gateway.
2. Disseminate relevant **information** by uploading it in the Contribution Gateway (News and Events, Case Studies, AI Assets, Open Calls, Educational resources…).
3. Visit other projects' profiles and use the functionality "**follow project**" to receive their newsletter and be up-to-date on their progress and achievements.
4. Use the "**Newsletter Builder**": this tool enables the AIoD users managing European projects to build Newsletters for their projects and distribute them via email to other members of the Platform "following" the project.
5. Use the "**Website builder**", a open-source licensed Word Press theme for creating websites of European-funded projects that are associated with the AIoD. The website has "connectors" to link specific content to the AIoD platform.

Funded by the European Union

Contacts

Funded by
the European Union

Contact details

To ask any general questions or doubts about the platform	info@aiod.eu
To send images and contents to be communicated live on social media	comms@aiod.eu

Funded by
the European Union

AI on Demand

Knowledge and services
for the AI community

Thank you!

Funded by
the European Union

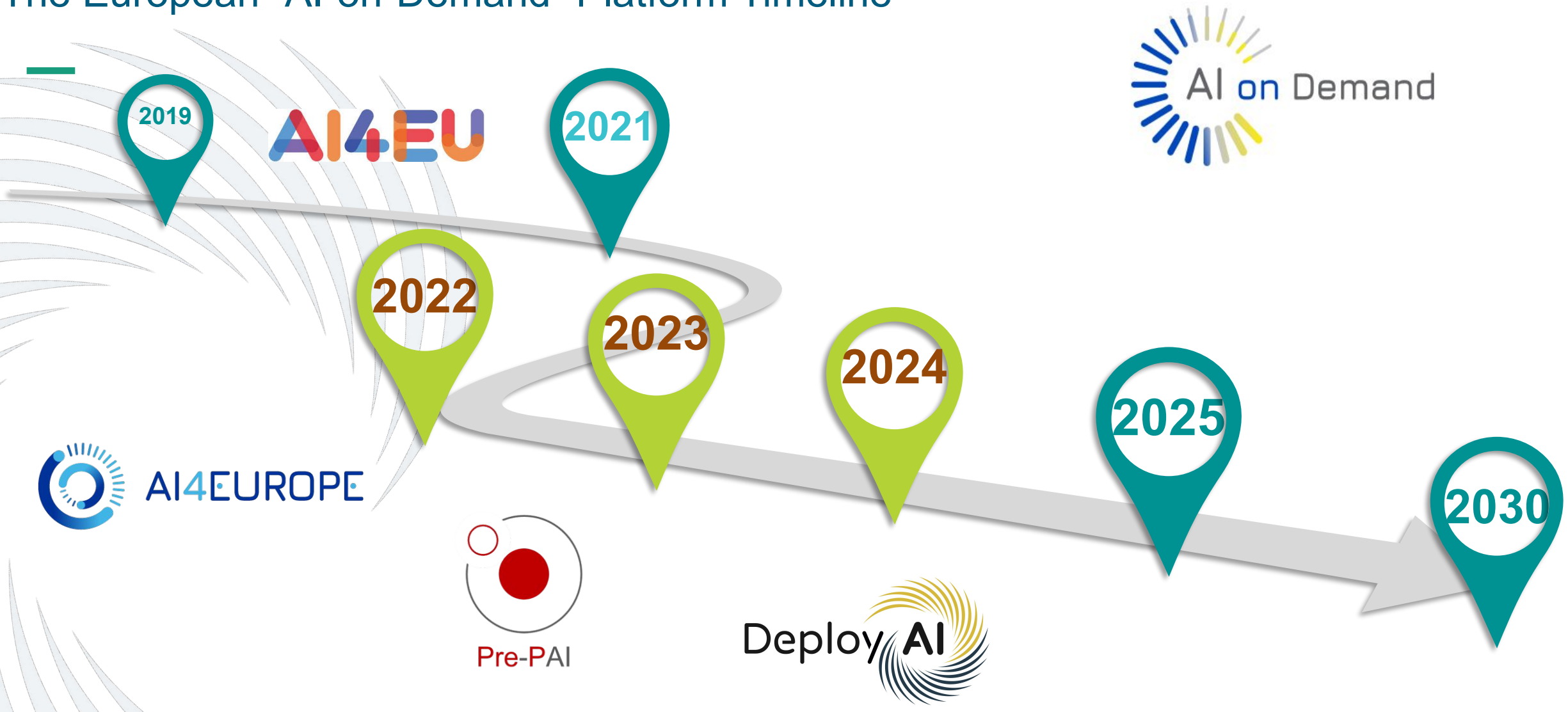
DeployAI

Development and Deployment of the European AI-on-demand Platform

Denia Kanellopoulou, NCSR “Demokritos”

DG-CONNECT Webinar: EDIHs - AI-on-Demand Platform | 24.04.2024

The European “AI-on-Demand” Platform Timeline



*(*provisional logo)*

The Goal of DeployAI project

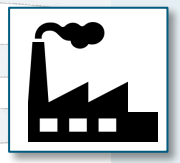
Deploy, and launch a fully operational AI-On-Demand platform (AIoDP)

building on the work of AI4Europe and

promoting **trustworthy, ethical, and transparent European AI solutions**

for use in the **industry** and in the **public sector**.

Strategic Perspective of DeployAI project (I)

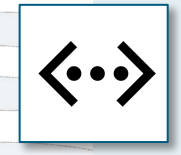


Industrial and professional platform capabilities (TRL 9)

Attractive for European SMEs, start-ups and public administration beyond the EU project ecosystem

Stakeholder-driven development engaging feedback and requirements of SMEs and large Industries, public bodies, TEFs & (E)DIHs

Strategic Perspective of the DeployAI project (II)



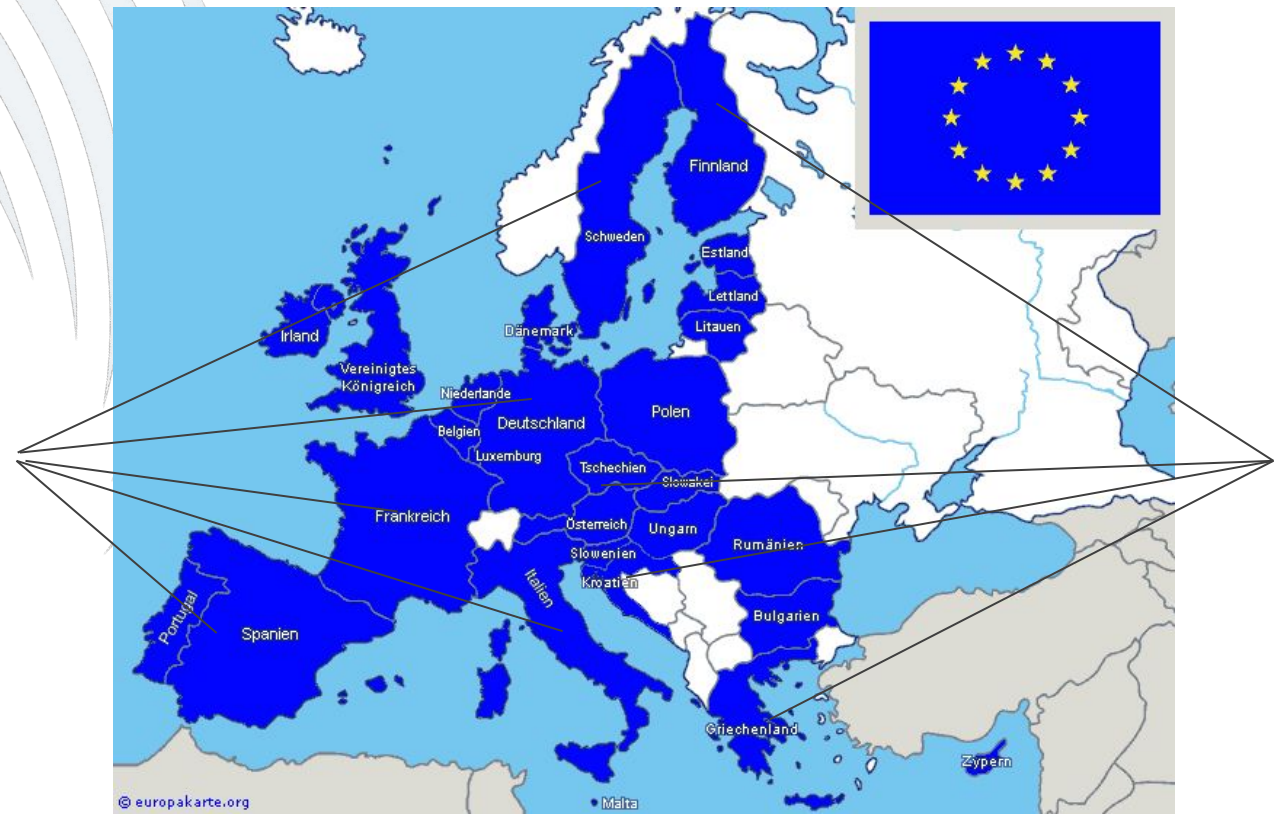
Interoperability with European Data Spaces, Gaia-X, TEFs, EDIHs, HPC systems (including EuroHPC), EOSC, ELG, European Cloud & Edge services, and industrial AI-capable Cloud platforms

Open Source and Openness

Services for Generative AI and HPC

Top Tier European Companies & Institutions working together

28 European Partner | 13 European Countries



A strategic mix of partners for a strong consortium that will deliver value

- Strong AI competences
- Ready-to-use critical infrastructures (HPC, cloud-to-edge, dataspaces, LLMs)
- Strong involvement in AIoD activities: PrePAI and AI4Europe
- Active in EuroCC, EDIHs, TEFs, Dataspaces and other flagship EU initiatives
- Strong collaborations with other key stakeholders in the EU community
- Industrial IT companies with experience in software and platform development

Input from PrePAI Stakeholder Consultations: What do (E)DIHs expect from the European AI-on-demand Platform

Pre-PAI

- Connection and networking opportunities with the **AI ecosystem, TEFs and Data Spaces**
- Access to reliable **data**
- Secure access to **HPC, cloud and edge infrastructures**
- **Training** tools, including peer-learning and training on trustworthy AI and AI regulations
- Support in the development of an **AI adoption roadmap** for SMEs
- Support in the implementation of their **test-before-invest services**
- Ready-to-use **AI tools and resources** for practical applications
- **Trustworthiness** assessment of their AI solutions
- Access to success stories, best practices, quality assurance methodologies and guidelines

Let’s discuss on how DeployAI can best serve the EDIHs and their customers needs!

Thank you!

Denia Kanellopoulou
denia@iit.demokritos.gr

AI on Demand

Knowledge and services for the AI community

The Metadata Catalogue - REST API

EDIHs webinar

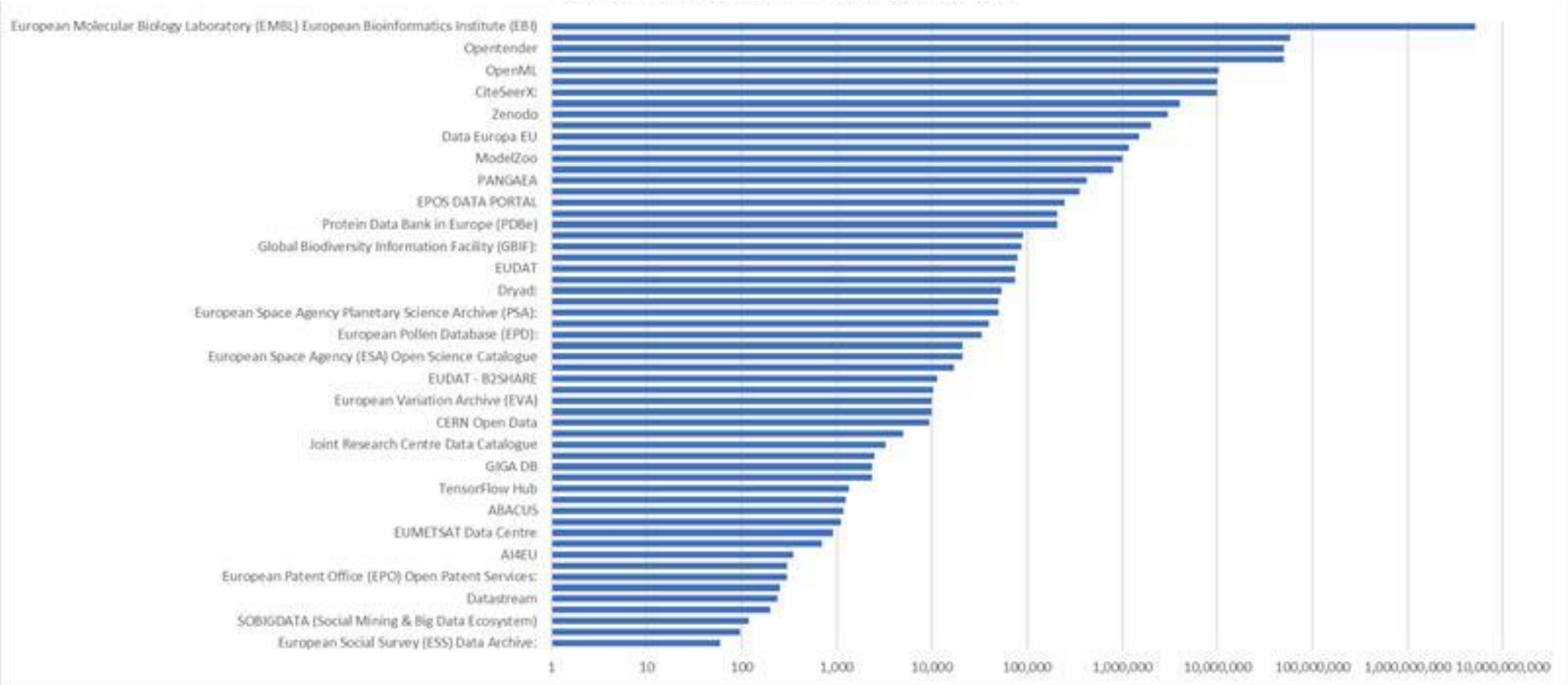
24-Apr-2024

Jean Matias

University College Cork (UCC)

Funded by the European Union

Number of Assets in Data and AI Portals



Funded by
the European Union

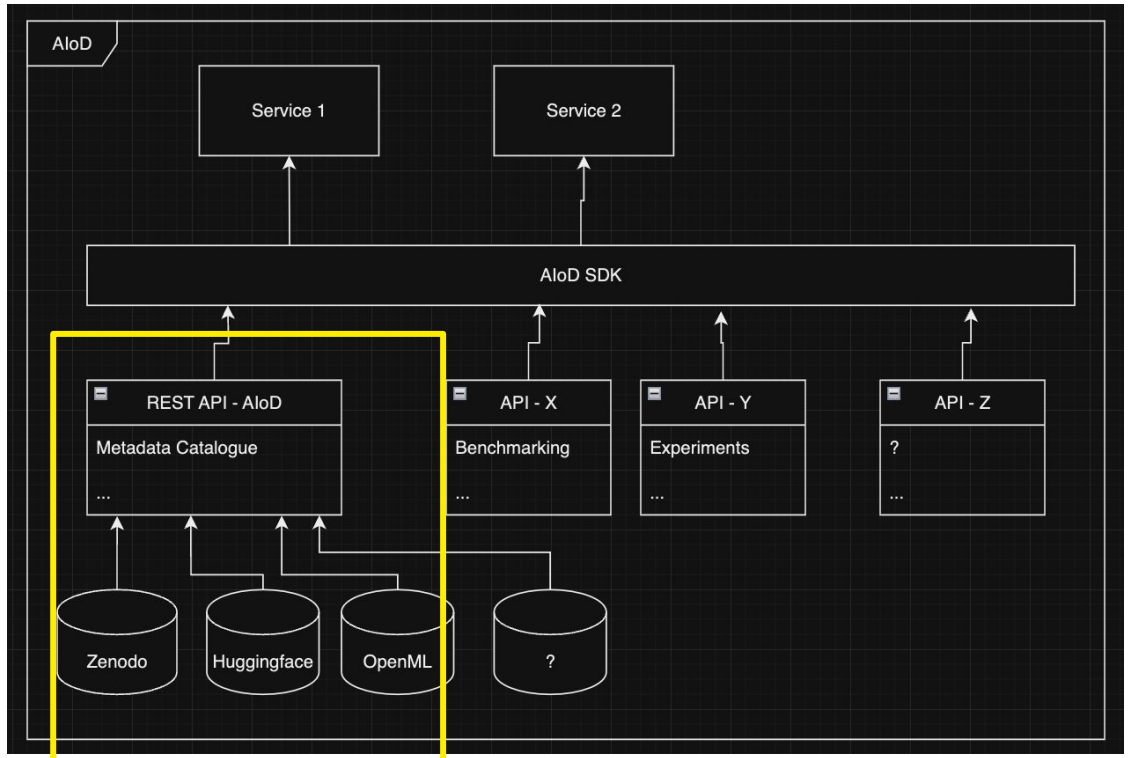
The AIoD Metadata Catalogue

The challenge

⚠️ Spread across several portals/platforms
⚠️ No standard data/metadata models
⚠️ Lack of quality standards/assessments
⚠️ No extraction standard protocol

The Solution

✅ Single interface
✅ Multiple connectors to external platforms
✅ Single protocol for download and upload
✅ Unified metadata model
✅ Shareable with the AI community
⬜ Metadata/Data quality and assessments



12K+ **ML Models**

400K+ **Datasets**

Distributed REST API
Containerized

Funded by the European Union

The AIoD Metadata Catalogue

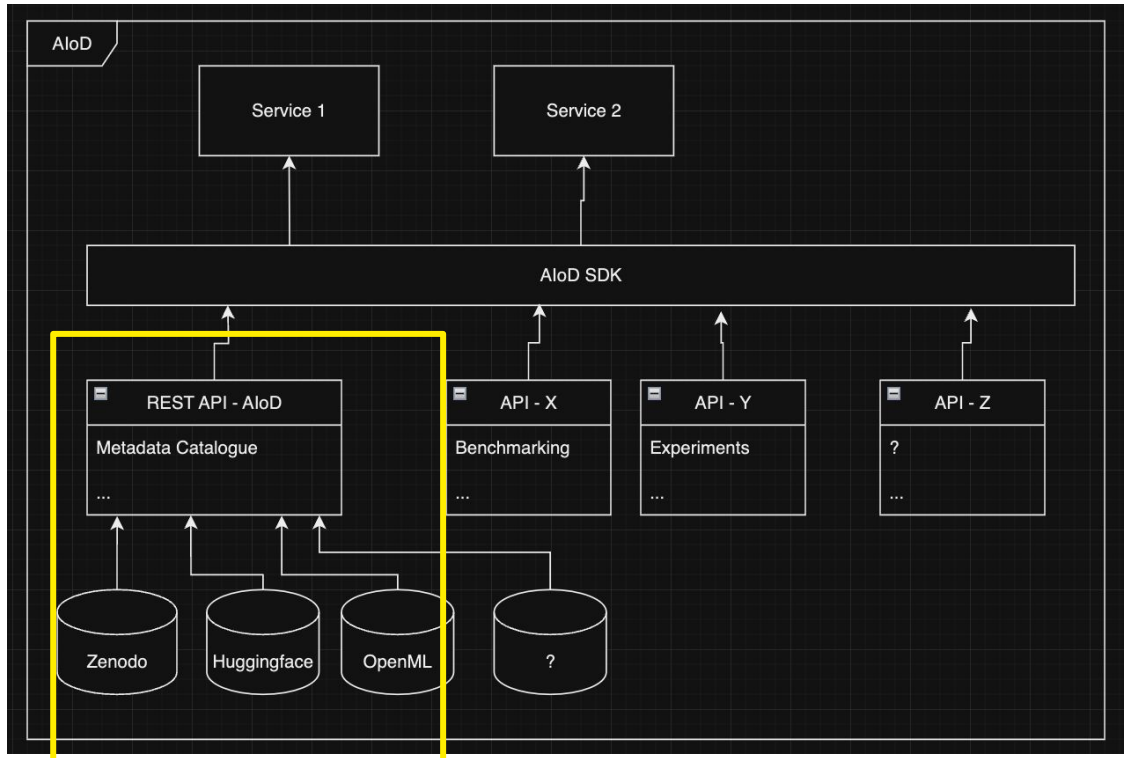
Whom is this for?

✅ Researchers (PhD, MSc Students, etc.)

1. To train AI and/or ML models
2. To test existing AI and/or ML models
3. Share assets (Datasets, Models, etc.)

✅ Developers

1. Who want to use AIoD platform as a backend for services (Websites, Applications, etc.)



12K+ **ML Models**

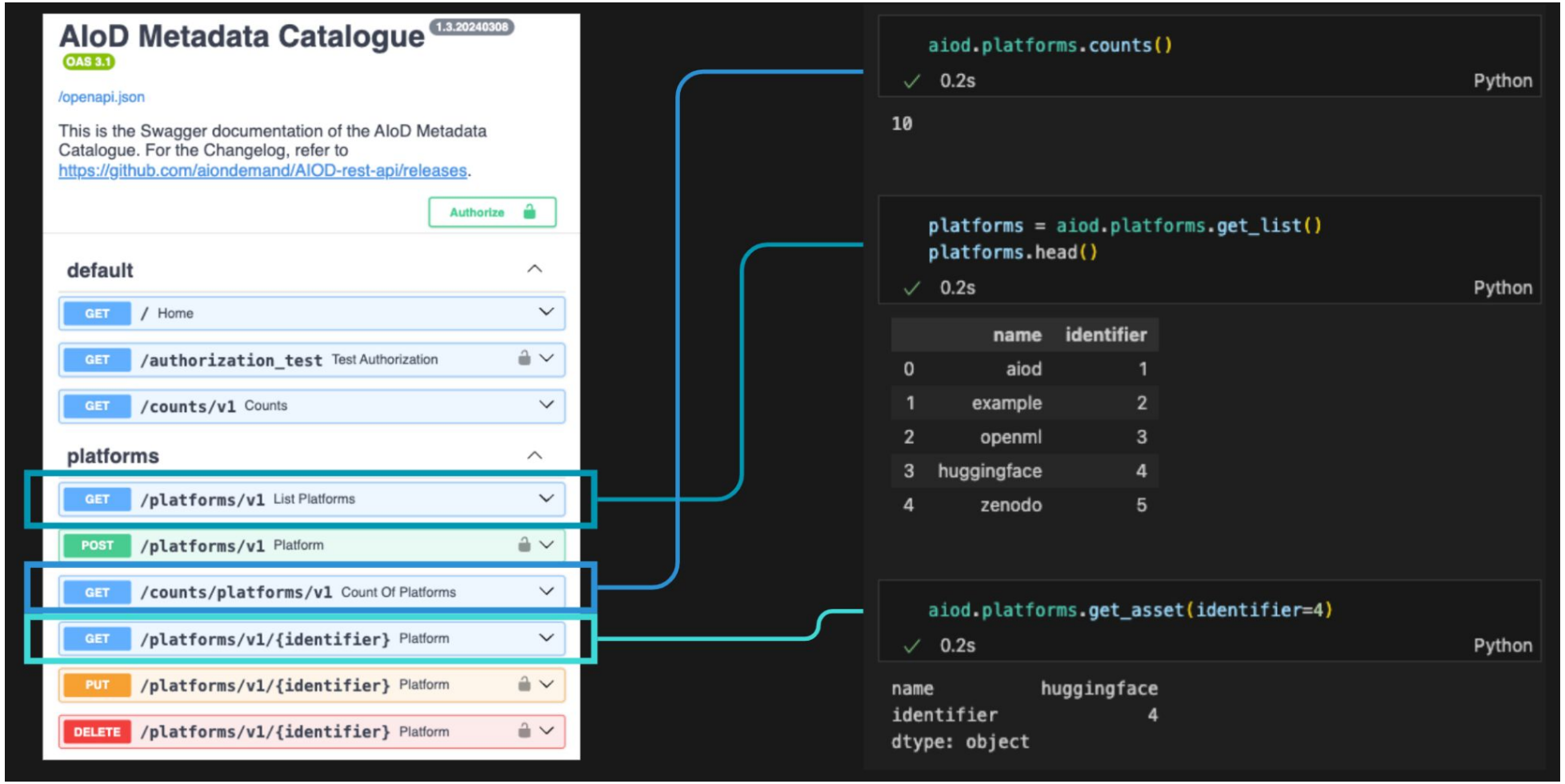
400K+ **Datasets**

Distributed REST API

Containerized

Funded by the European Union

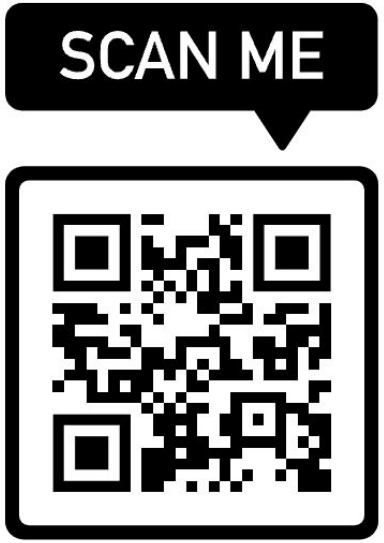
The AIoD Metadata Catalogue



Funded by
the European Union

A walkthrough and Hands-on with the AIOD SDK

https://aiod.eu/



Next steps…

- ⬜ Add connectors
- ⬜ Improve the search engine
- ⬜ Integrate features for reproducibility
- ⬜ Integrate features for experimentations
- ⬜ etc…

Funded by
the European Union

AI on Demand

Knowledge and services
for the AI community

Thank you!

SCAN ME

Funded by
the European Union

AI on Demand

Knowledge and services for the AI community

AIoD from the perspective of a service RAIL: Research and Innovation AI Lab

Martin Tamajka, KInIT

2024-04-24, EDIHs - AI on Demand Platform

Funded by the European Union

Outline

- AIoD from the perspective of a service
  - How can a service integrate and benefit from AIoD?

- RAIL: Research and Innovation AI Lab
  - An example of a particular service integrated with AIoD

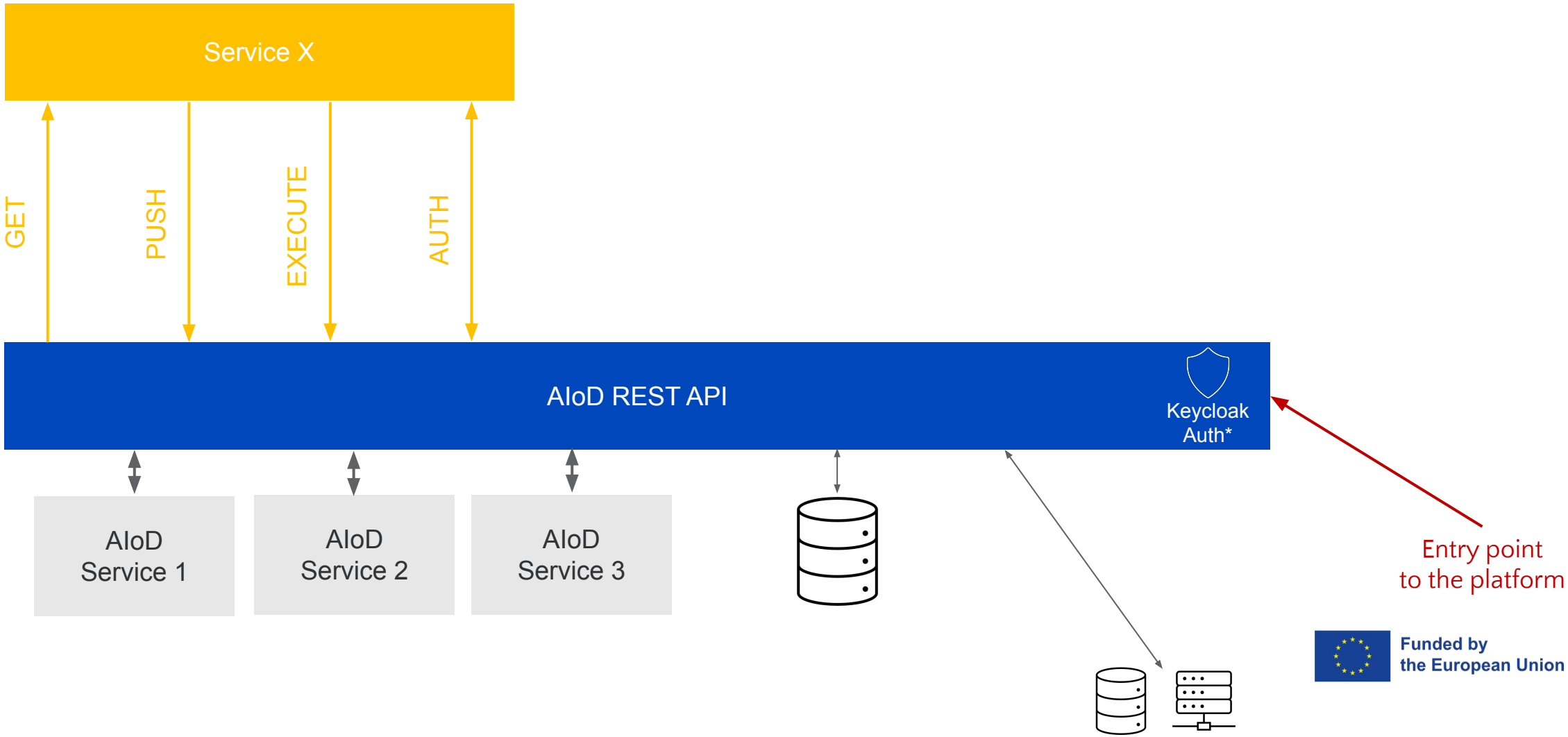
Funded by
the European Union

AI on Demand platform

… from the perspective of a service

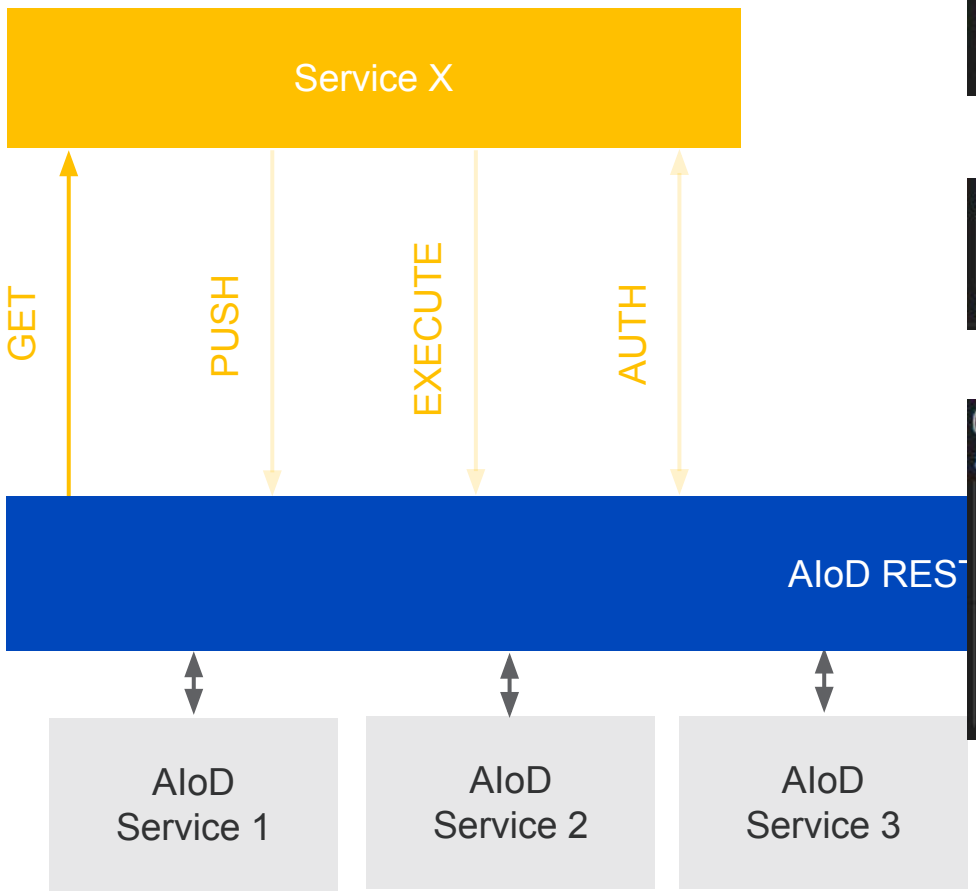
Funded by
the European Union

Interacting with the platform means interacting with the REST API (simplified view)



Funded by
the European Union

Interacting with the platform means interacting with the REST API (simplified view)



Javascript

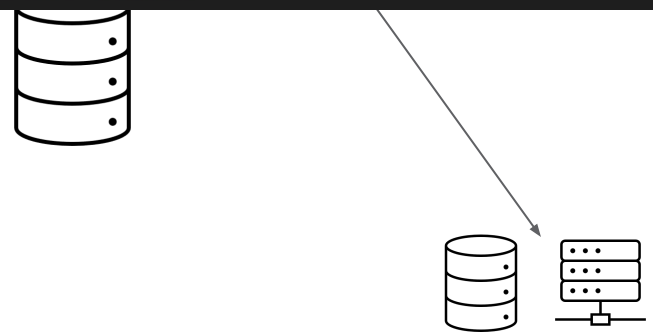
```javascript
let datasets = await fetch('http://localhost:8001/datasets/v0?schema=aiod&offset=0&limit=5')
    .then(datasets => datasets.json());
```

Typescript (Angular 16)

```typescript
getDatasets(offset: number = 0, limit: number = 10): Observable<Dataset[]> {
  return this.http.get<Dataset[]>(`${environment.BACKEND_API_URL}/assets/datasets?offset=${offset}`
  + `&limit=${limit}`);
}
```

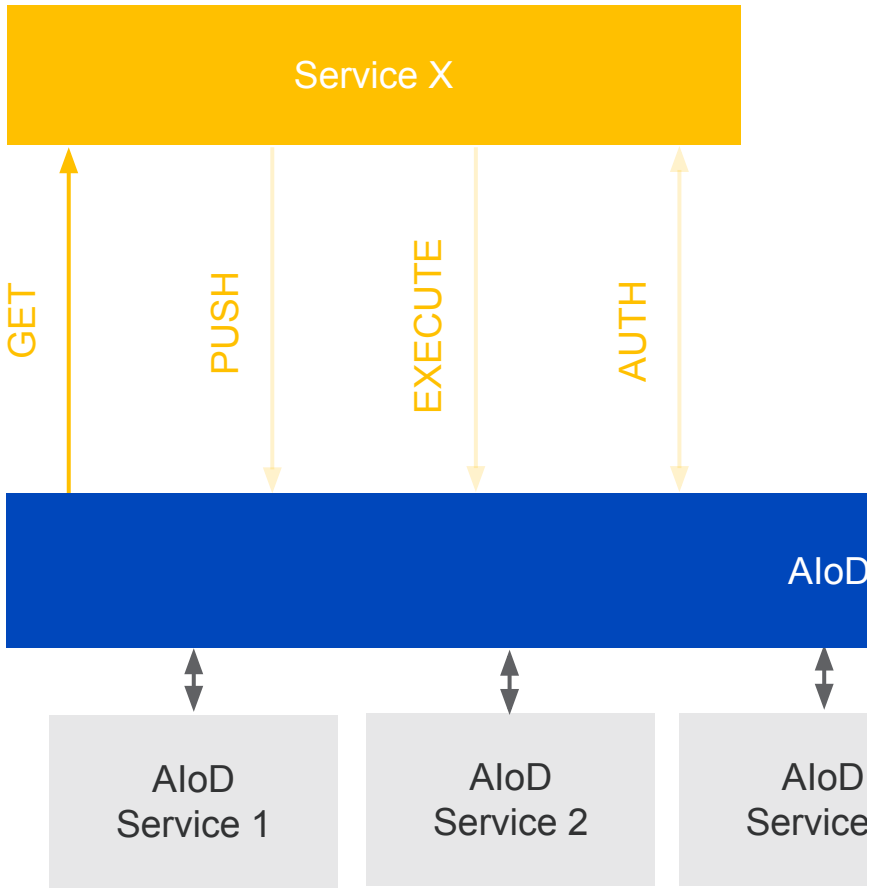
Python (FastAPI)

```python
@router.get("/datasets", response_model=list[Dataset])
async def get_datasets(pagination: Pagination = Depends()) -> Any:
    async_client = aiod_client_wrapper()
    res = await async_client.get(
        f"{settings.AIOD_API.BASE_URL}/datasets/{settings.AIOD_API.DATASETS_VERSION}",
        params={"offset": pagination.offset, "limit": pagination.limit},
    )
    return res.json()
```



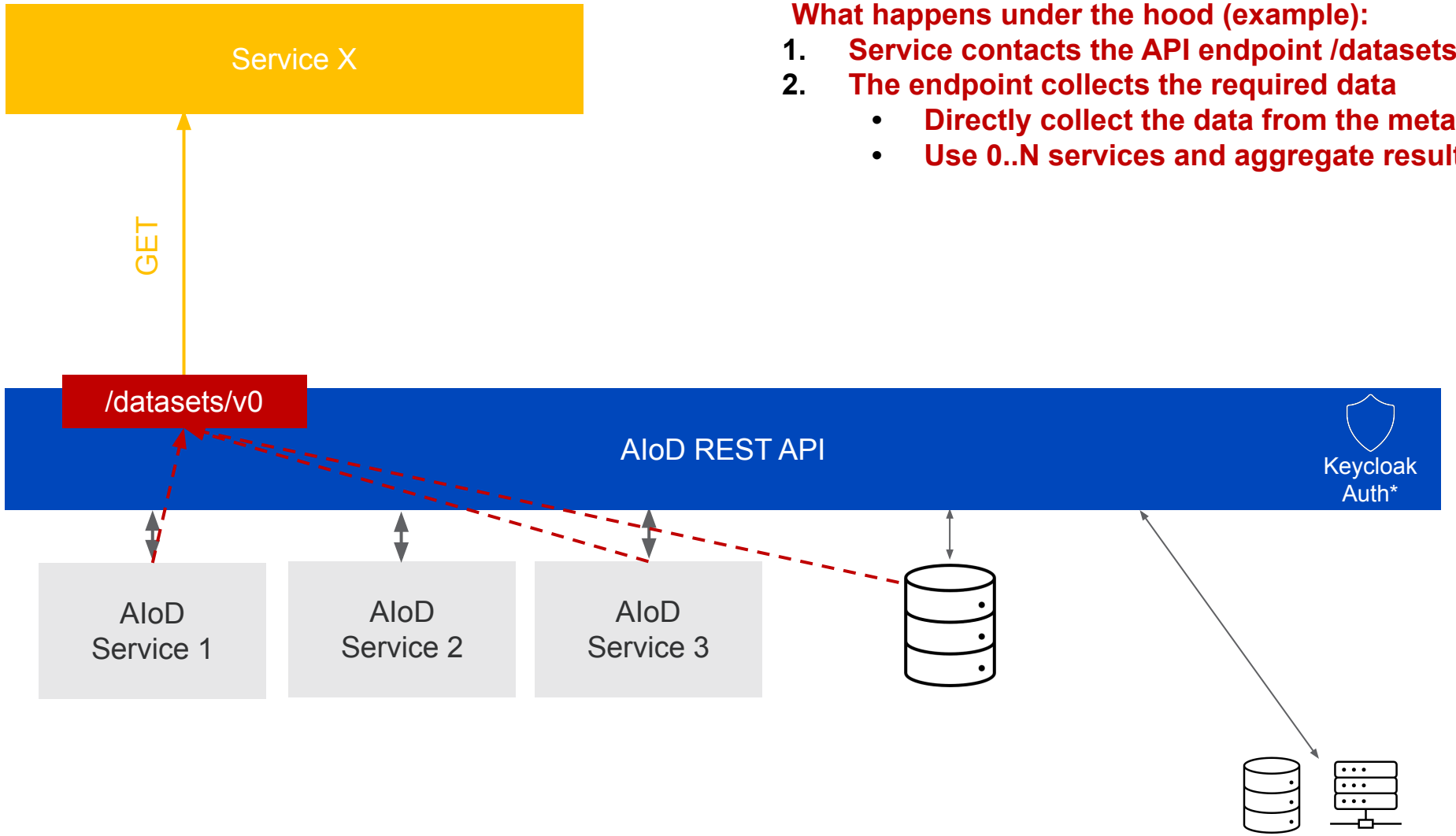
Funded by the European Union

Interacting with the | with the REST API (si




```
[
  {
    "platform": "openml",
    "platform_identifier": "2",
    "description": "**Author**: Unknown. Donated by David Sterling and Wray Buntine  \n\n**Source**: [UCI](https://archive.ics.uci.edu/ml/datasets/Annealing) - 1990  \n\n**Please cite**: [UCI](https://archive.ics.uci.edu/ml/citation_policy.html)  \n\n\n\nThe original Annealing dataset from UCI. The exact meaning of the features and classes is largely unknown. Annealing, in metallurgy and materials science, is a heat treatment that alters the physical and sometimes chemical properties of a material to increase its ductility and reduce its hardness, making it more workable. It involves heating a material to above its recrystallization temperature, maintaining a suitable temperature, and then cooling. (Wikipedia)\n\n\n\n### Attribute Information:\n\n     1. family:  --,GB,GK,GS,TN,ZA,ZF,ZH,ZM,ZS\n\n     2. product-type:    C, H, G\n\n     3. steel:           -,R,A,U,K,M,S,W,V\n\n     4. carbon:          continuous\n\n     5. hardness:        continuous\n\n     6. temper_rolling:  -,T\n\n     7. condition:       -,S,A,X\n\n     8. formability:     -,1,2,3,4,5\n\n     9. strength:        continuous\n\n    10. non-ageing:      -,N\n\n    11. surface-finish:  P,M,-\n\n    12. surface-quality: -,D,E,F,G\n\n    13. enamelability:   -,1,2,3,4,5\n\n    14. bc:              Y,-\n\n    15. bf:              Y,-\n\n    16. bt:              Y,-\n\n    17. bw/me:  B,M,-\n\n    18. bl:              Y,-\n\n    19. m:               Y,-\n\n    20. chrom:           C,-\n\n    21. phos:            P,-\n\n    22. cbond:  Y,-\n\n    23. marvi:           Y,-\n\n    24. exptl:           Y,-\n\n    25. ferro:           Y,-\n\n    26. corr:            Y,-\n\n    27. blue/bright/varn/clean:          B,R,V,C,-\n\n    28. lustre:          Y,-\n\n    29. jurofm:          Y,-\n\n    30. s:               Y,-\n\n    31. p:  Y,-\n\n    32. shape:           COIL, SHEET\n\n    33. thick:           continuous\n\n    34. width:           continuous\n\n    35. len:  continuous\n\n    36. oil:             -,Y,N\n\n    37. bore:            0000,0500,0600,0760\n\n    38. packing: -,1,2,3\n\n    classes:  1,2,3,4,5,U\n\n  \n\n    -- The '-' values are actually 'not_applicable' values rather than\n\n       'missing_values' (and so can be treated as legal discrete\n\n       values rather than as showing the absence of a discrete value).",
    "name": "anneal",
    "same_as": "https://www.openml.org/api/v1/json/data/2",
    "date_modified": "2019-07-09T15:22:03",
    "date_published": "2014-04-06T23:19:24",
    "is_accessible_for_free": true,
    "size": 898,
    "version": "1",
    "alternate_names": [],
    "citations": [],
    "distributions": [
      {
        "content_url": "https://api.openml.org/data/v1/download/1666876/anneal.arff",
        "encoding_format": "ARFF",
        "checksum": []
      }
    ],
    "is_part": [],
    "has_parts": [],
    "license": "Public",
    "keywords": [
      "study_1",
      "study_76",
      "study_37",
      "uci",
      "study_34",
      "study_70",
      "study_14",
      "test",
      "study_41"
    ],
    "measured_values": [],
    "identifier": 1
  },
  { … }, // 19 items
  { … }, // 19 items
  { … }, // 19 items
  { … } // 19 items
```

Interacting with the platform means interacting with the REST API (simplified view)

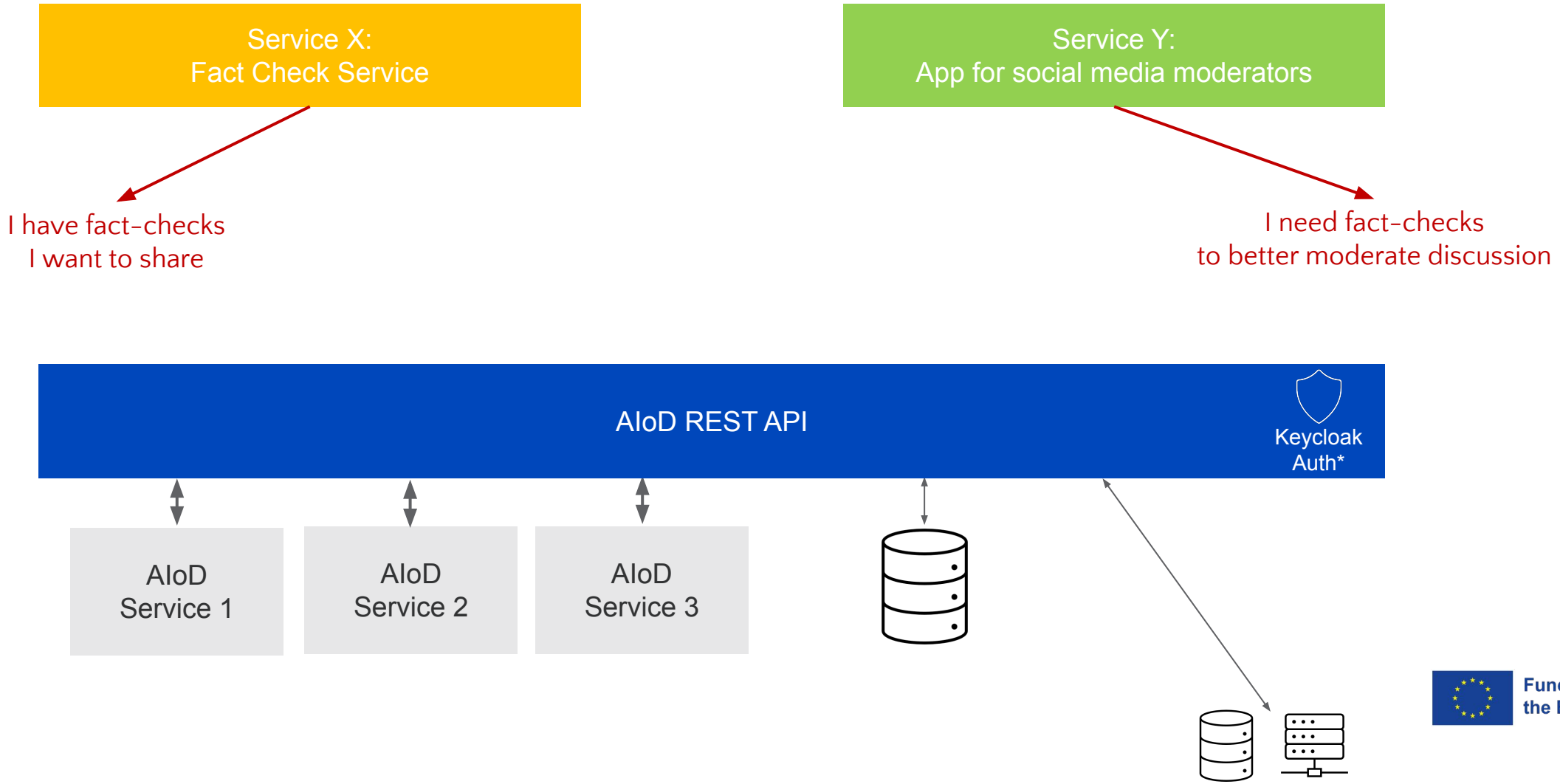


What happens under the hood (example):

1. **Service contacts the API endpoint /datasets/v0?schema=aiod**
2. **The endpoint collects the required data**
   - **Directly collect the data from the metadata store.**
   - **Use 0..N services and aggregate results.**

Funded by
the European Union

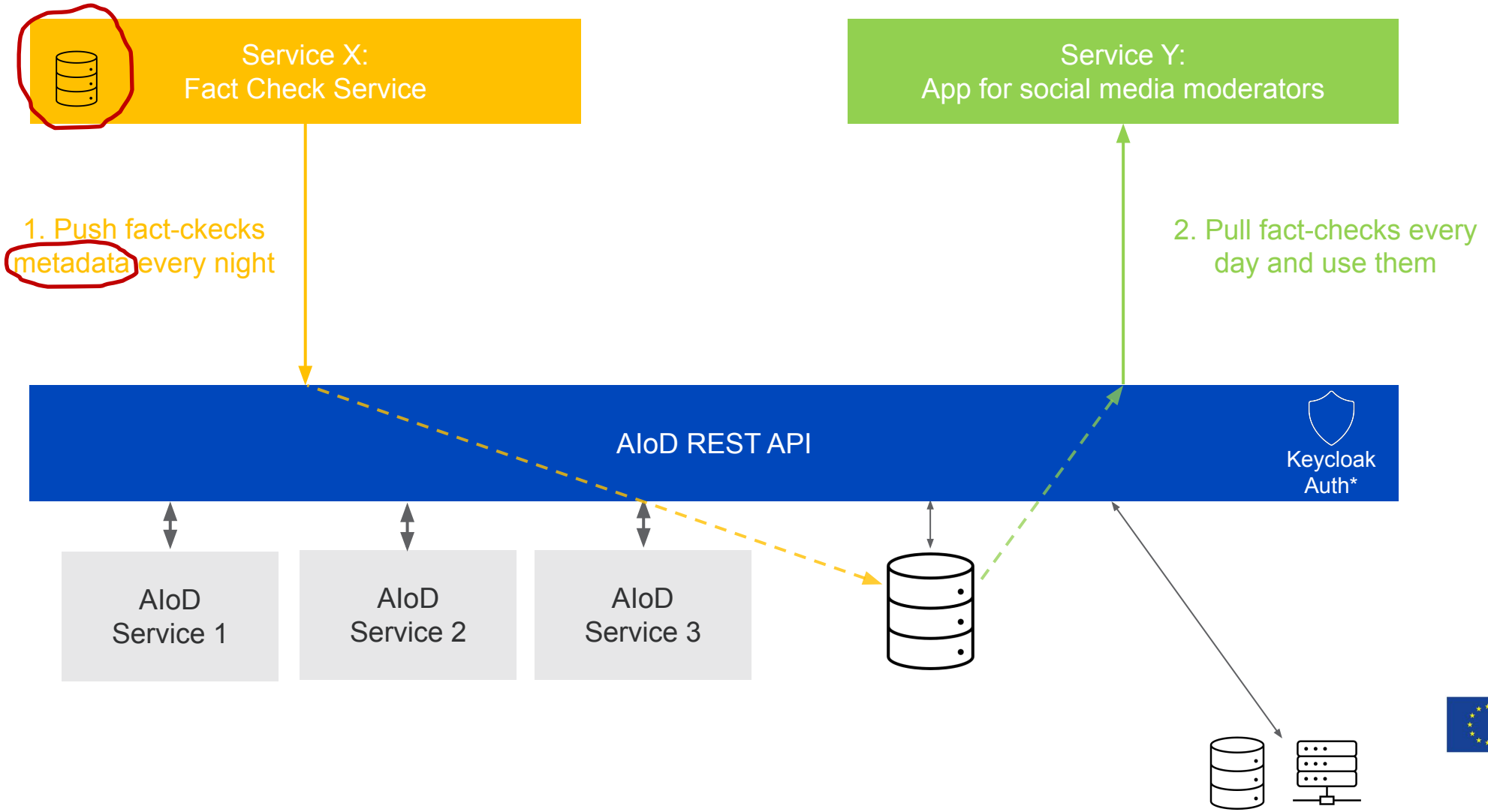
Services can benefit from each other through AIoD (without being aware of one another)



Funded by
the European Union

Services can benefit from each other through AIoD (without being aware of one another)

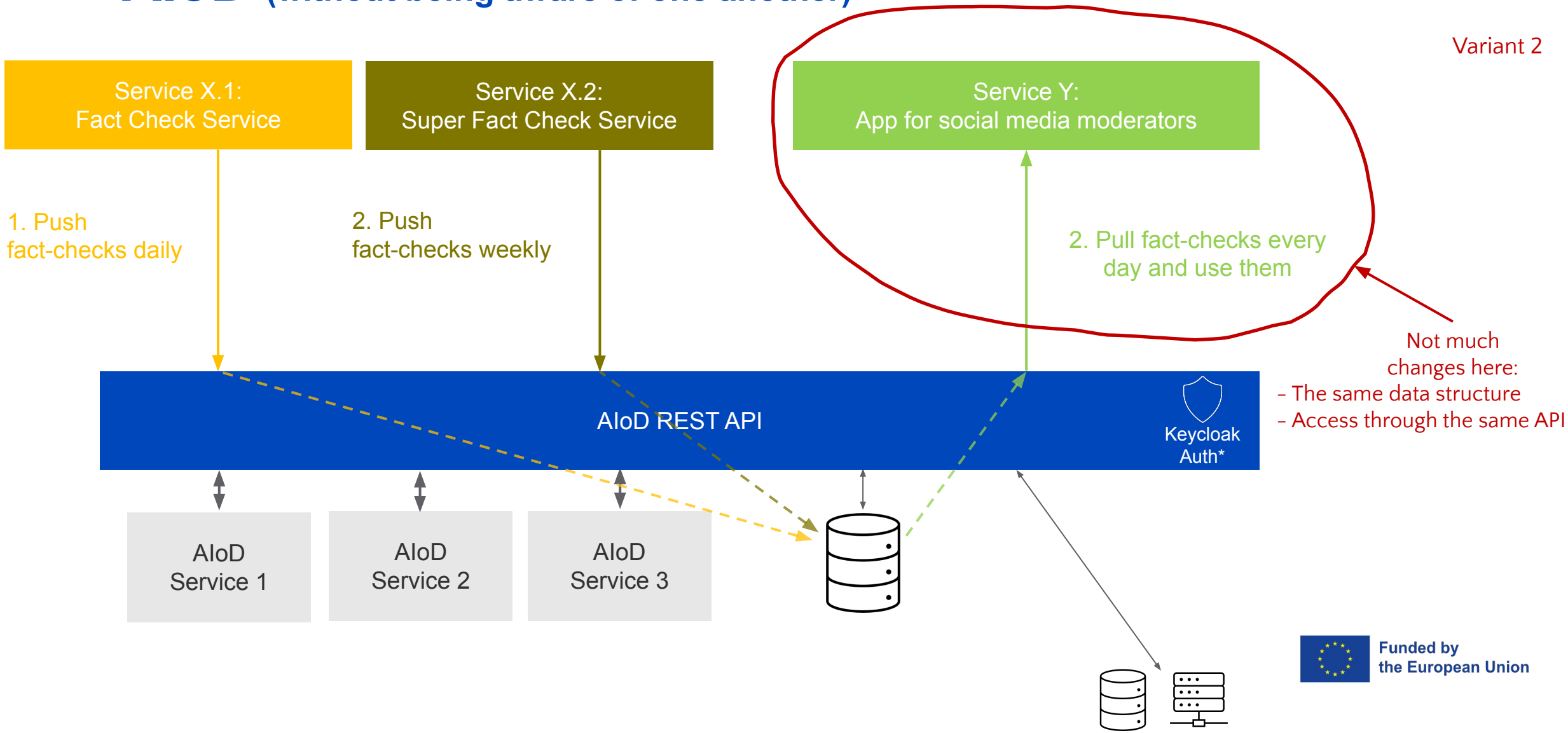
Variant 1



Funded by the European Union

Services can benefit from each other through AIoD (without being aware of one another)

Variant 2



Funded by
the European Union

RAIL: Research and Innovation AI Lab

… one particular service integrated with AIoD

Funded by
the European Union

RAIL is:

Simple

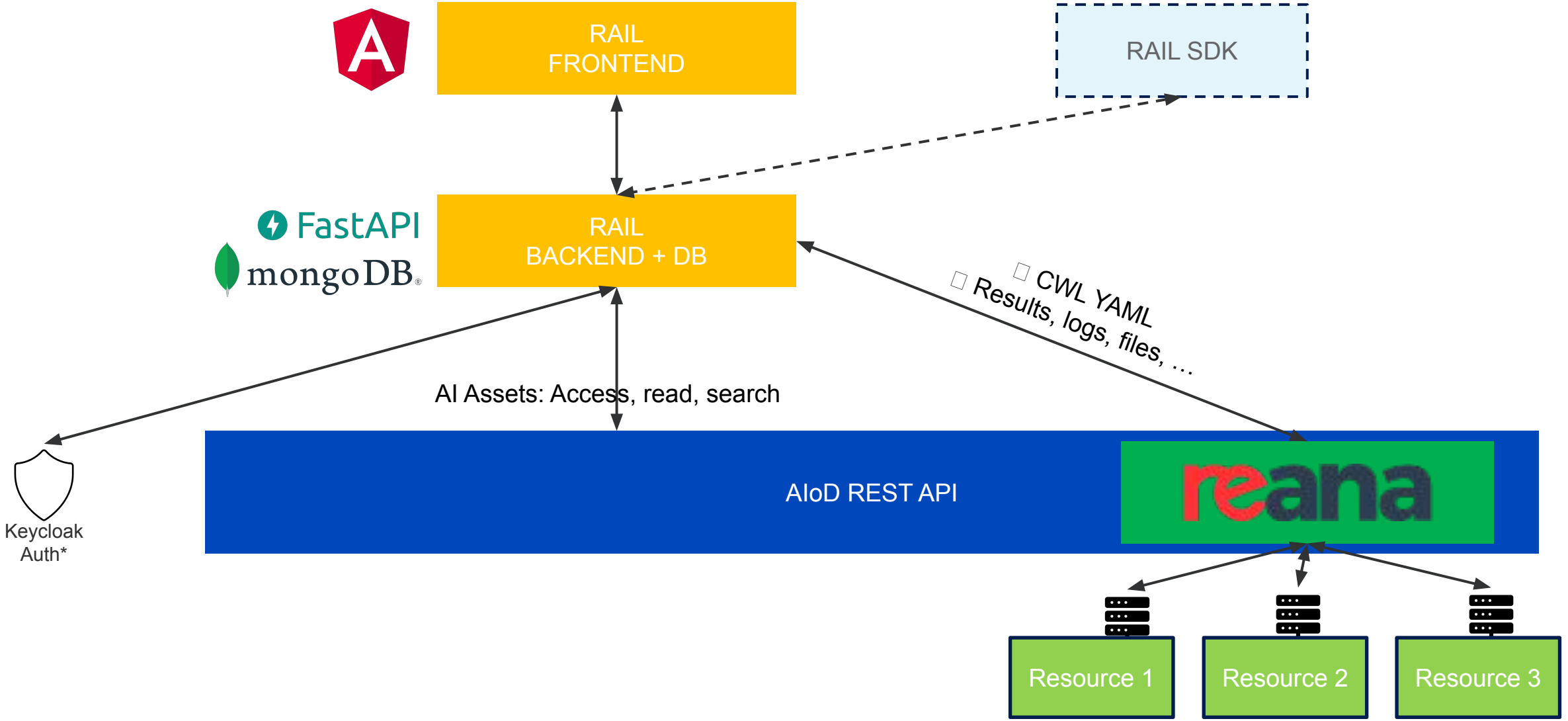
- A tool that allows AI practitioners to explore and use AI assets directly in AIoD … in a lightweight and flexible way

More extensive

- A **web application** (currently also working on API and SDK)
- … that enables **AI practitioners**
- … to **work with AIoD AI Assets** (explore, search, compare, …)
- … and **create experiments** that are reproducible and reusable
- … that are **executable directly in the AIoD** platform supported by its infrastructure
- … and **download the results** of the experiments.

Funded by the European Union

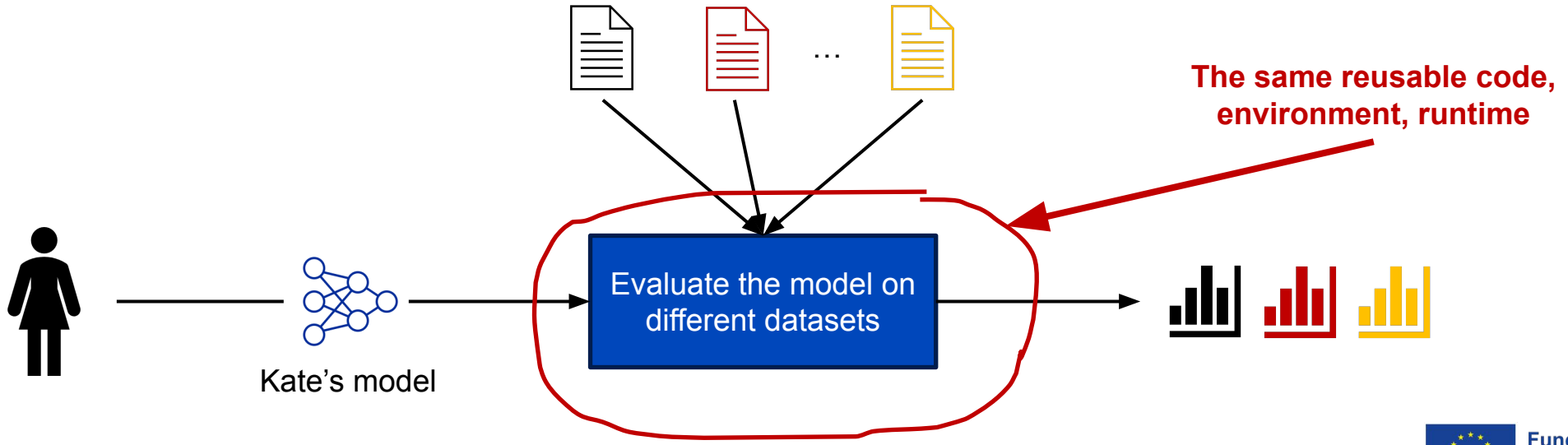
What's under the hood



RAIL enables you to create and execute experiments … that you and others can modify and reuse

Let’s illustrate RAIL on a simple use case:

- Kate is a CTO of a small startup that needs to check how well her sentiment analysis model performs in different scenarios (different datasets)



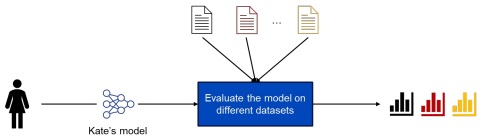
Funded by the European Union

How can RAIL and AIoD help?

The metadata about models for sentiment analysis is already there

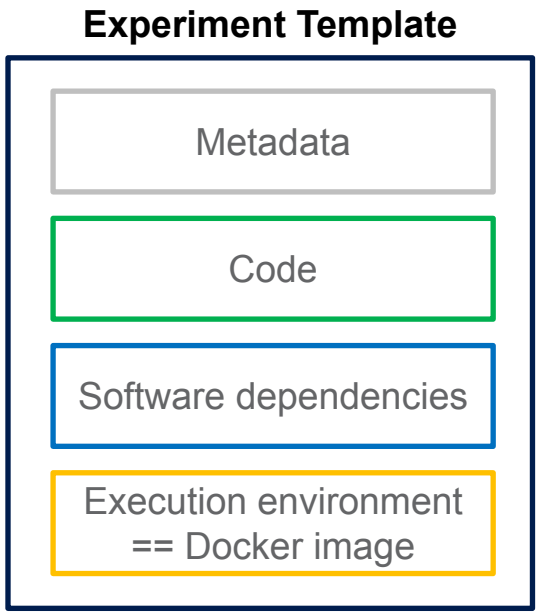
Two scenarios

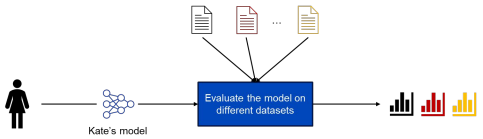
1. Someone already implemented reusable “Sentiment evaluation experiment” in RAIL and Kate can reuse it
2. Kate implements her own “Sentiment evaluation experiment” and uses it for multiple datasets



Funded by the European Union

Create and execute an experiment
3 main components of experiment in RAIL





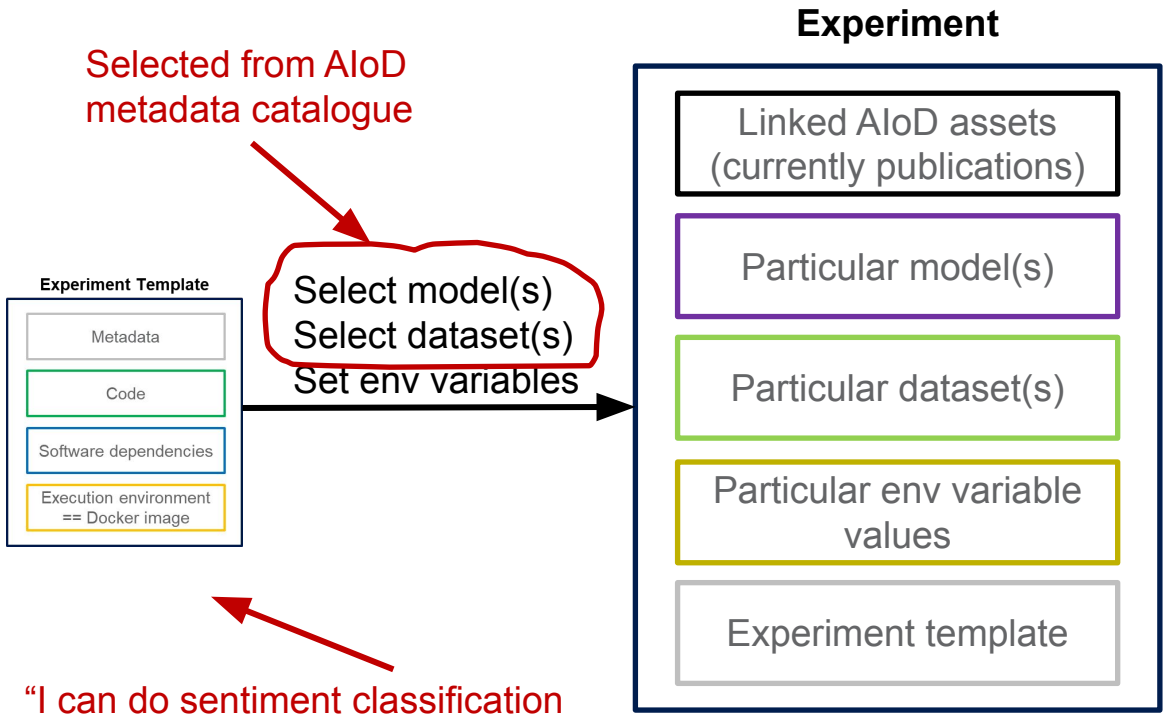
Demo: Create Experiment template

- Under the hood (after approval):
  - Build a Docker image and install dependencies
  - Pack in the code
  - Push the image to a registry

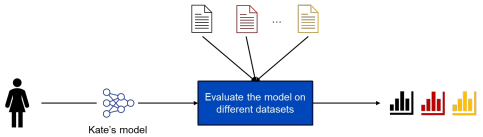
Experiment template

- On its own it's not executable
- You need to specify particular parameters the template should be executed with
  - Model, environment variables, …
  - These are processed by the code

Create and execute an experiment
3 main components of experiment in RAIL



"I can do sentiment classification with HuggingFace model and HuggingFace dataset, but you need to tell me:
- which model should I use
- which dataset should I use
- which training split should I use
- which column in the dataset should I use."



Demo: Create Experiment

- Under the hood:
  - A record in a MongoDB is created
  - That's it


```
_id: ObjectId('6627782429c2c3f4476682fb')
name: "Sentiment-IMDB"
description: "dsa"
▸ publication_ids: Array
updated_at: 2024-04-23T08:58:12.230+00:00
created_at: 2024-04-23T08:58:12.230+00:00
created_by: "martin.tamajka@kinit.sk"
experiment_template_id: ObjectId('6627750029c2c3f4476682fa')
▾ dataset_ids: Array
    0: 288
▾ model_ids: Array
    0: 2
▾ env_vars: Array
  ▾ 0: Object
      key: "SPLIT"
      value: "train"
  ▾ 1: Object
      key: "TEXT_COLUMN"
      value: "text"
  ▾ 2: Object
      key: "HOW_MANY_SAMPLES_TO_PROCESS"
      value: "100"
```

Create and execute an experiment
Experiment in RAIL

Demo: Run an experiment and collect results

- Under the hood:
  - Create a CWL manifest
    - "Inject" datasets, models as environment variables for the script
      - Now, they're translated to model name and dataset name
  - Send CWL to AIoD instance of REANA
    - Wait until the experiment is executed
  - Show logs, enable downloading results (model, data, …)

RAIL can…

- Utilize AIoD API
  - Datasets, models, (publications)
- Run jobs on the AIoD computational infrastructure
- Compose assets into reusable executable experiments
- **To be deployed in next two weeks**
  - Download results of these experiments
  - Utilize assets I “buy” in another AIoD service - MyLibrary

- Nice example of a reusable pipeline
  - Someone creates an experiment to fine-tune an LLM for low resource languages (a lot of work)
  - Ideally, someone just injects his/her own data and reuses the whole pipeline directly in AIoD and downloads the model

Funded by
the European Union

(some) Future directions

- Use RAIL directly in Python / Jupyter notebooks
  - SDK (Python)
    - ```
      import aiod
      dataset = aiod.get_dataset(id=XYZ);
      model = aiod.get_model(id=ABC)
      experiment = aiod.create_experiment(dataset, model, id=PQR)
      result = aiod.execute(experiment, infrastructure=JKL)
      ```
  - Inner and outer SDK
- Execute code from public GitHub repository
  - Plus, a browser extension
- Compose experiments to pipelines
  - Select dataset =[exp1: translate]=> Select model =[exp2: predict sentiment]=> Results
- Benchmarks, leaderboards, …
  - Helpful visualization to compare my experiments (e.g., performance of different models on my dataset) and also public experiments Community effect
- Support for other languages than Python
      ```

AI on Demand

Knowledge and services
for the AI community

Thank you!

Funded by
the European Union

AI on Demand

Knowledge and services for the AI community

AI-Builder

EDIHs webinar
24-Apr-2024

Sangamithra Panneer Selvam
Fraunhofer IAIS

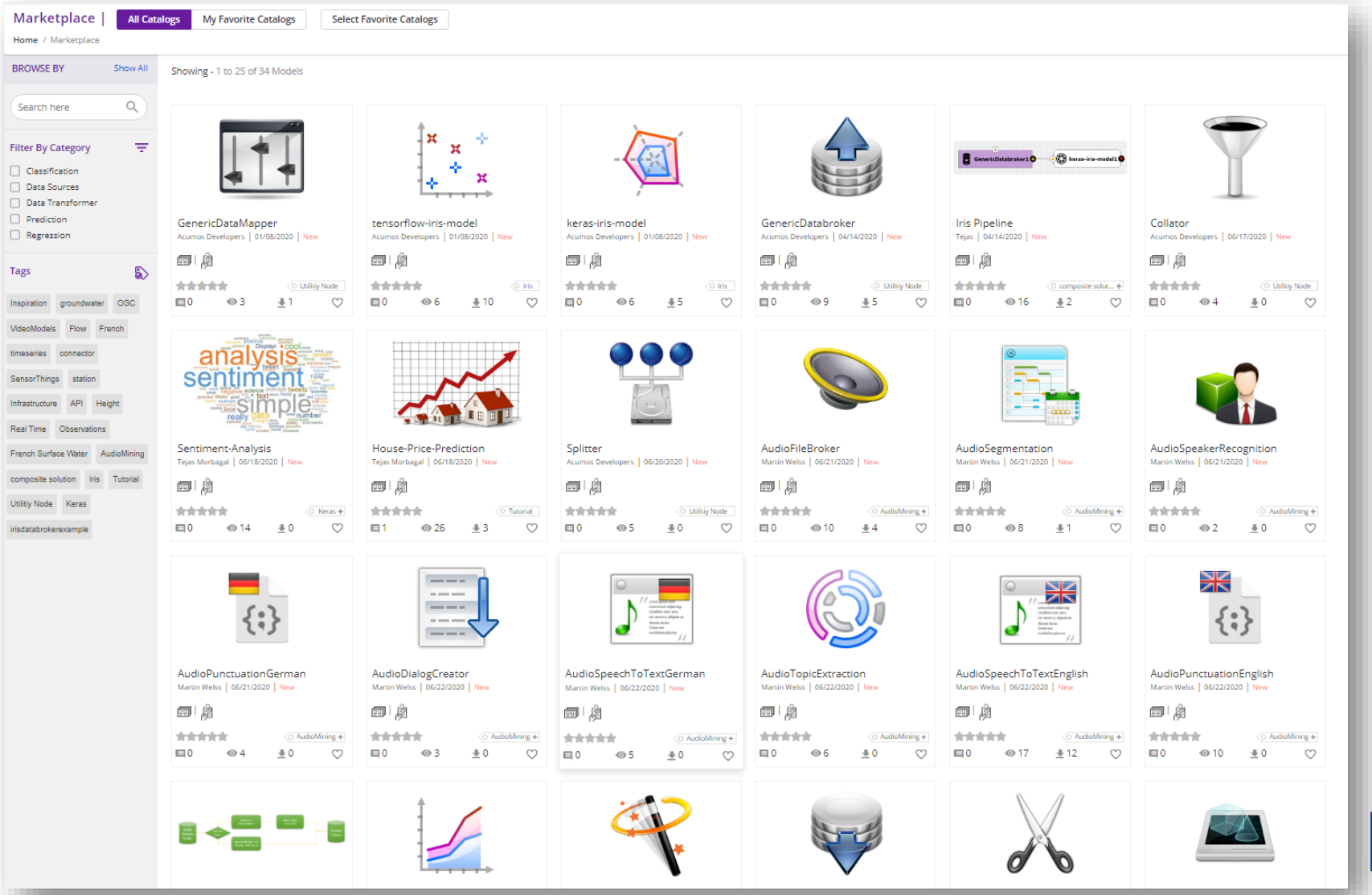
Funded by the European Union

AI-Builder Objectives

AI-Builder is a framework to compose and deploy cognitive architectures from re-usable AI-Modules.

Funded by
the European Union

Catalog of modular, re-usable AI tools

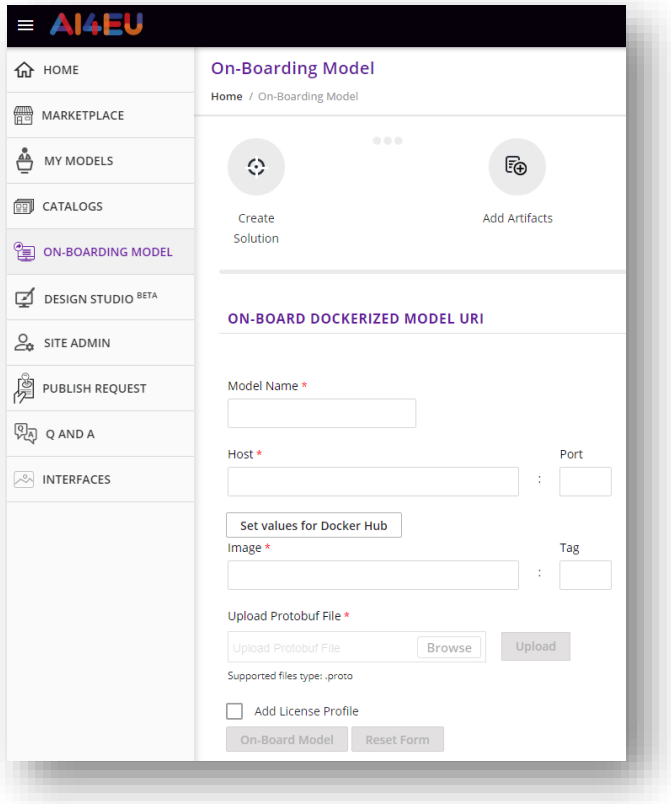


Funded by
the European Union

Onboarding and Publication

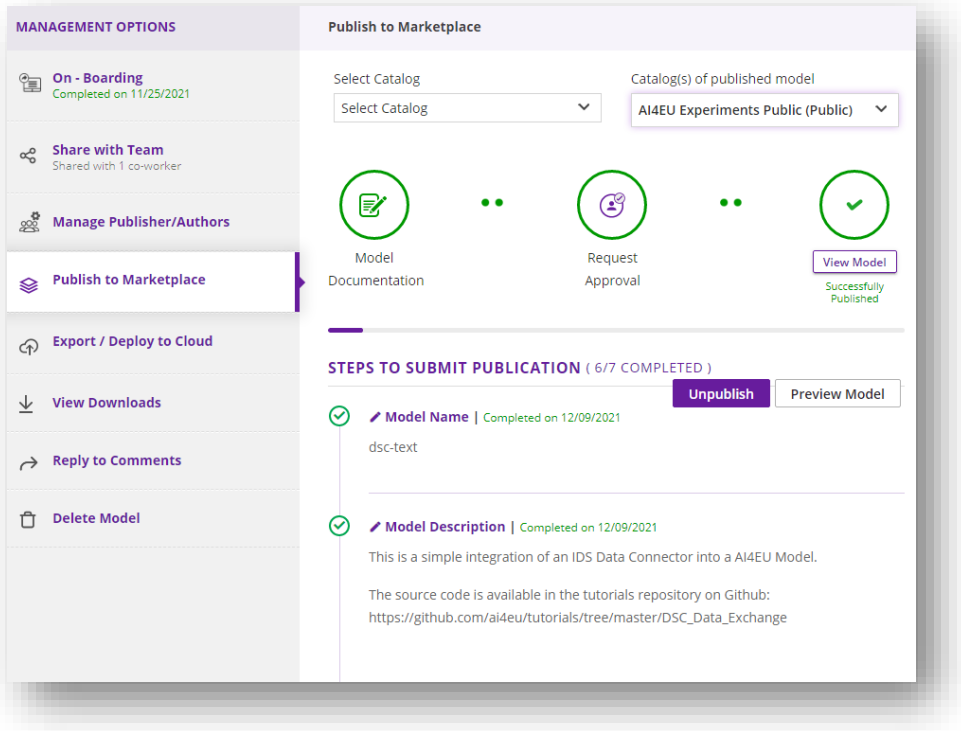
To add a tool to the catalog, two steps are necessary:

1. Onboarding



[Onboarding Tutorial]

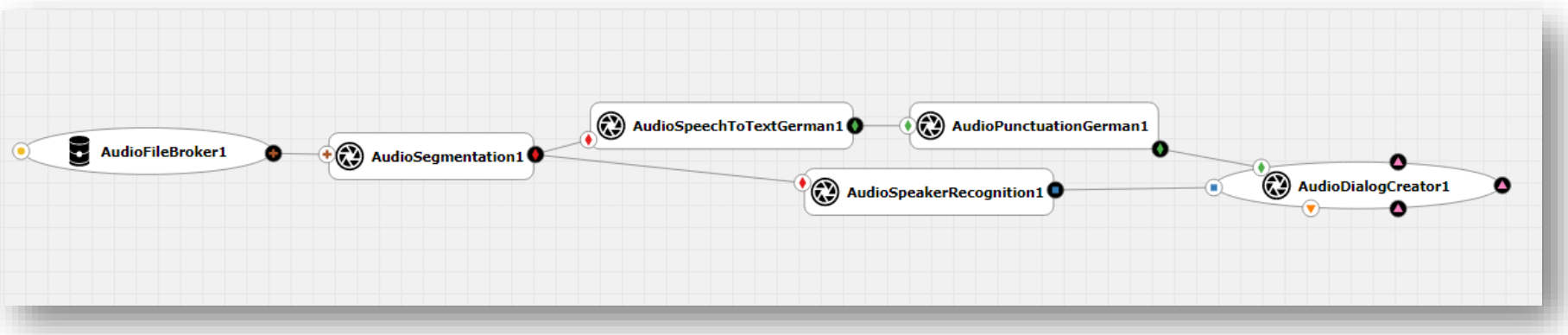
2. Publication



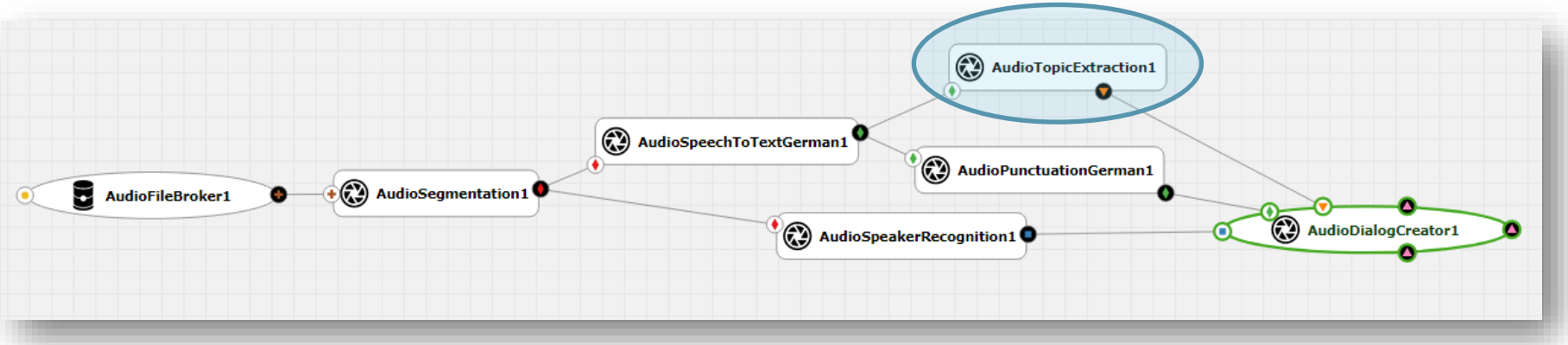
[Publication Tutorial]

Funded by the European Union

Visual AI Pipeline Composition



- Audio Pipelines composed of re-usable building blocks
- Visual Editor checks for compatible interfaces
- Pipelines start with a databroker
- Let non AI experts create and deploy pipelines



Funded by
the European Union

AI4EU Container Specification

- Docker container
- Protobuf specification of public interface
- gRPC communication
- optional Web-UI for human interaction
- Based on free / open source technologies
- Recommendations for scalability, training and GPU-Support
- Support for gRPC streaming

```
// set used version of protobuf
syntax = "proto3";



// define input data structure
message IrisDataFrame {
  repeated double sepal_length = 1;
  repeated double sepal_width = 2;
  repeated double petal_length = 3;
  repeated double petal_width = 4;
}

// define output data structure
message ClassifyOut {
  repeated int64 value = 1;
}

// define exposed service
service Model {
  rpc classify (IrisDataFrame) returns (ClassifyOut);
}
```

Funded by
the European Union

Important Links

- AI4Europe Drupal Portal: https://www.ai4europe.eu/
- AI-Builder + AI Playground: https://aiexp.ai4europe.eu/#/home
- Container Specification: https://gitlab.eclipse.org/eclipse/graphene/tutorials/-/tree/main/Container_Specification
- Source Code AI-Builder: https://gitlab.eclipse.org/eclipse/graphene
- Graphene Tutorials: https://gitlab.eclipse.org/eclipse/graphene/tutorials
- YouTube Playlist: https://www.youtube.com/playlist?list=PLL80pOdPsmF6s6P6i2vZNoJ2G0cccwTPa

Funded by
the European Union

AI on Demand

Knowledge and services for the AI community

Thank you!

www.aiod.eu info@aiod.eu @AIonDemand

Funded by the European Union