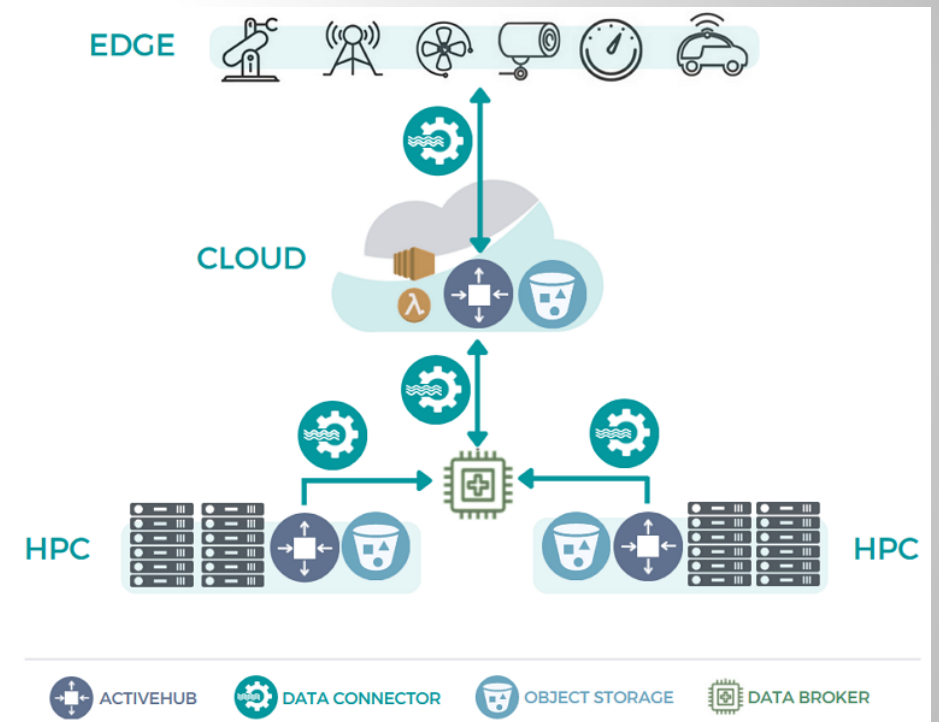# What is NEARDATA project?

NEARDATA is an open platform for data scientist to help them to **process and analyze large scientific data** in different formats such as text, images, and videos.

It offers solutions both in the **Cloud and on devices like IoT (Edge) and high-performance computers (HPC)**.

NEARDATA uses advanced **AI to make data analysis faster and smarter**, while also adding a strong layer of security to ensure that **data is kept safe and private**.

EDGE

CLOUD

HPC   HPC

ACTIVEHUB   DATA CONNECTOR   OBJECT STORAGE   DATA BROKER

NEARDATA

NEARDATA
Scientific Fields

What fields does the project work on?

PROJECT FOCUS

NEARDATA is targeting health data in three key domains (project use cases):

- **Metabolomics** (images)

- **Genomics** (text)

- **Surgery** (video)

Each facing significant data volume and processing challenges.

NEARDATA

NEARDATA Challenges

What problems does it address?

KEY ISSUES

Unstructured Data

80-90% of data is unstructured (e.g., text, images, video), making it difficult to analyze due to its lack of searchable formats. Managing this data remains a major hurdle for data scientists.

Data Volume

Large health datasets push technology limits, making data analysis expensive and complex.

Real-Time Data Processing Needs

Surgery data requires fast video analytics, bringing data processing close to where it is obtained (Edge).

Privacy & Security

Sensitive health data imposes strict confidentiality, making data sharing and extraction difficult.
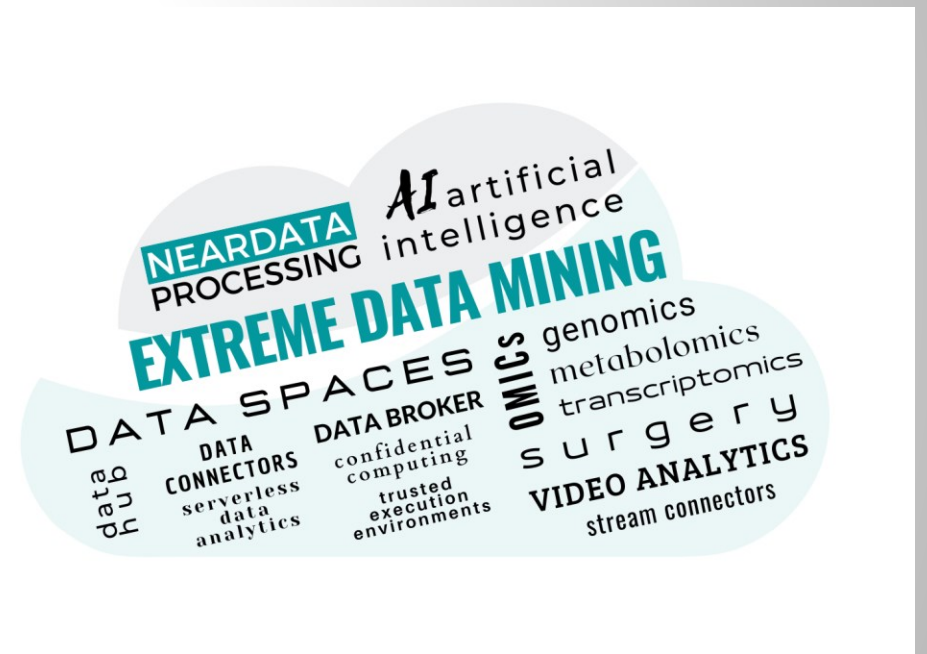
NEARDATA

NEARDATA Objectives

O1 To build a powerful platform that **improves the connection of data between storage and analytics platforms**.

O2 To support fast **data processing in real-time** for video analytics.

O3 To create a confidential layer that will **ensure secure access** to private data sources by using advanced security technologies.

NEARDATA

NEARDATA is working on three health data domains:

Metabolomics
Expand the analysis of open or sensitive metabolomics data in the METASPACE platform to ensure the external access and efficient re-use of data within the platform.

Genomics
Creation of functionalities that allow to analyze large volume of genomics data efficiently in the cloud or on the edge.

Surgery
Create AI models that can aid surgeons during surgery and allow fast analysis of video data in real-time.

Use cases

Use case is a practical application focused on solving a specific problem in the field scientific research

NEARDATA

Use cases benefits

The use cases have achieved the following benefits in managing extreme data:

Metabolomics

Thanks to the NEARDATA project, METASPACE (EMBL) has been able to process larger metabolomics datasets faster using new AI technologies that help the scientific community to find and develop new medicines in the fields of biology, medicine and pharmacology.

Genomics

With UKHS and SANO, we have been able to interpret massive data that have improved how different hospitals respond to health emergencies.

With BSC, we have focused on helping hospitals better understand and address rare diseases through new data analysis techniques.

Surgery

With NCT, we have developed fast and safe technologies that make assisted and robotic surgery possible. Surgeons can benefit from these to facilitate and improve surgeries.

NEARDATA

Key Technologies

Lithops, METASPACE, SCONE, and Pravega
are mature platforms supporting the project, with more than three years of live, active communities, and created and controlled by project partners.

Partners

NEARDATA

Communication and dissemination activities

Achievements in communication during the first half of the project

Type of communication	Category of audience	Achievements
Scientific Publications	Scientific community	17 publications
Conferences and Workshops	Scientific community and Industry	44 events
Community Building	Scientific community, Industry	4 events
Meetings outside the consortium	Scientific community, Industry	23 meetings
Events for society	General Public	29 events
Press releases	General Public	6 publications

NEARDATA