



MIMER

Storskaligt användande av AI

med fokus på tillgängliggörande av metadata
som du inte visste att du hade

Joel Hedlund, Data Science Director

NAISS



RI
SE



Joel Hedlund, Data Science Director

National Academic Infrastructure for
Supercomputing in Sweden ([NAISS](#))

Analytical Imaging Diagnostic Arena ([AIDA](#))



MIMER

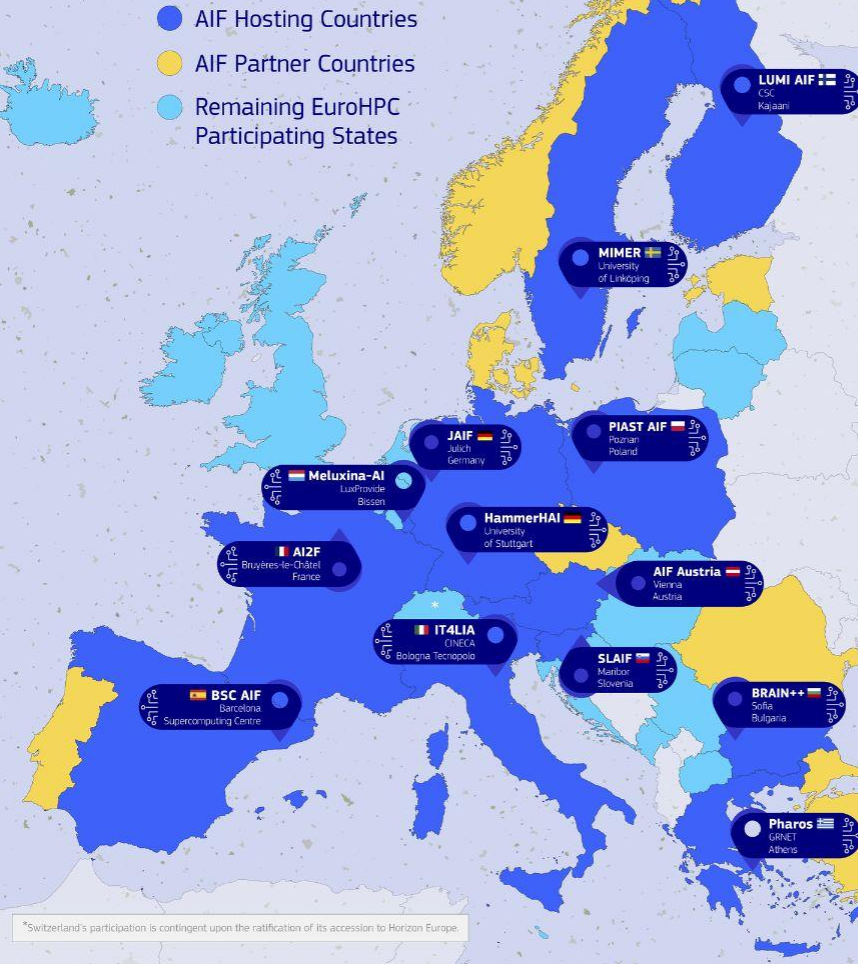
The Mimer AI Innovation Factory

NAISS



**RI.
SE**

AI Factories



2+ BEUR European Commission Initiative

13 AI Factories across Europe managed by EuroHPC JU

Facilitating AI innovation for SMEs and the public sector



Mimer supercomputer

A pure AI system.

Sensitive data ready.

Strong customer separation.

Customer rules, customer accountability.

Free access for selected excellent projects,
including SMEs under de minimis rules.

Open to further customers.

Expect delivery Jul 2026, operational Oct.

Graphic from AIDA Data Hub [Data Science Platform](#), with permission.



Mimer offering

Dynamic network of AI experts: 52 FTE.

Full AI chain support: develop, test, deploy.

Hands-on support and training.

Two (or more) co-working hubs.

Hosting and sharing of data and models.

Co-located with national flagship resources.



NAISS Arrhenius

Next flagship resource for academic computing in Sweden, part of EuroHPC. Largest ever in Sweden.

Being installed as we speak.

10% of budget to sensitive data support.



What do I get as a customer?

Assistance in getting free access to compute and storage, exempt from state aid rules.

Support and training (state aid rules, under de minimis regulation):

- Onboarding.
- Data management.
- Model architecture and optimisation.
- Integrated AI and domain expertise.
- Networking and co-working hubs
- Thematic workshops



**RI.
SE**



NAISS

Target Audiences

Startups, SMEs, Industry.

Healthcare & public administration.

Academic research.

Expect: ~1000 innovation projects served.

RISE Research Institutes of Sweden.

NAISS National Academic Infrastructure for
Supercomputing in Sweden.



Focus areas

Autonomous systems.

Gaming.

Life science.

Materials science.

(all domains are in scope)



European AI factory data labs

For interoperability and user/data mobility.

Mimer engagements:

Health and Life Science: **Coordinator**
BG, DE, ES, FI, FR, GR & LU contributing.

Public Administration: Contributor

Scientific Data EOSC: Contributor

May join other data labs at a later date.



Finansdepartementet

Försäkringskassan
Skatteverket

INKOM

Försäkringskassan
Hägersten

2025-07-11

Dnr.

Uppdrag till Försäkringskassan och Skatteverket att utreda
förutsättningarna för en AI-verkstad för den offentliga
förvaltningen

Regeringens beslut

Regeringen ger Försäkringskassan och Skatteverket i uppdrag att utreda
förutsättningarna för att etablera och förvalta en AI-verkstad – en
förvaltningsgemensam infrastruktur och tillhörande stödfunktioner för
utveckling och användning av artificiell intelligens (AI) i offentlig
förvaltning. När uppdraget utförs ska utgångspunkten vara att en AI-
verkstad, om den etableras, ska vara kostnadseffektiv, säker och robust. I
uppdraget ingår att lämna nödvändiga författningsförslag samt beskriva
eventuella konsekvenser för annan befintlig verksamhet inom Försäkrings-
kassan och Skatteverket.

Försäkringskassan och Skatteverket ska när uppdraget utförs föra en dialog
med Myndigheten för digital förvaltning (Digg), Kungl. biblioteket och
Riksarkivet samt med AI-noden Mimer vid Linköpings universitet.
Försäkringskassan och Skatteverket ska även inhämta synpunkter från
Upphandlingsmyndigheten, Kammarkollegiet, Konkursverket,
Integritetsskyddsmyndigheten (IMY) och Sveriges Kommuner och Regioner
(SKR) samt från andra relevanta aktörer.

Om Försäkringskassan och Skatteverket bedömer att det är nödvändigt för
att utföra uppdraget får myndigheterna genomföra försöksverksamhet i
begränsad omfattning.



Data Labs Services

Based on this analysis, Data Labs must provide a comprehensive set of services, including secure technical infrastructure for data transfer, storage, preparation, and synthetic data generation, along with compliant processing environments. They should also support data pooling through the discovery, acquisition, integration, and quality enhancement of datasets, while leveraging open data models.

Additionally, Data Labs must ensure regulatory compliance with EU legislation (such as GDPR, AI Act) by offering guidance, sector-specific support, and technical measures like anonymisation and secure environments to enable lawful and ethical AI development. (The identified services are listed in the annex.)

1. Implementation framework of Data Labs within AI factories

Data Labs will play a key role in making large volumes of data available to AI developers, the integration of multiple datasets—ensuring interoperability, high quality, and regulatory compliance. They will become cross-sectoral bridges to AI Factories, which are primarily aimed at operating within vertical data sectors (such as health and transportation).

AI factories operators are well positioned to deliver a set of Data Labs core services that address the essential needs of AI developers.

The list of proposed data labs Data Labs and participating AI factories:

1. Health and life science (coordinator SE, partners: FI, ES, LU, DE)–
2. Manufacturing (coordinator DE, partners FI, AT)
3. Education and culture (coordinator EL, partners: ES, DE, FI)
4. Public Administration (coordinator ES, partners: FI, DE, EL)
5. Cybersecurity (coordinator LU, partners: FI, DE, SE, EL)
6. Scientific Data EOSC (coordinator FI, partners: DE, SE, EL)
7. Destination Earth (coordinator ES, partners: FI, PL, LU)

AI Factory Operators have already identified a set of Data Labs services that will be delivered. These include:

Technical infrastructures and tools

- **Secure Data Transfer:** encryption protocols and secure communication channels, fast data transfer.
- **Secure and Safe Data Storage:** access controls, “large” storage capacity (e.g. hundreds of petabytes for cancer images), the data should have a storage time of decades for the important data. The systems should have measures in place to omit silent data corruption, even in case of long storage times.
- **Tools for data preparation for AI training:** data cleaning, normalization, and transformation tools.
- **Tools for creation of synthetic data:** computational time at AI factories to run algorithms and models (e.g. digital twins, LLMs) to generate artificial data.

- **Tools to ensure privacy-preserving processing of sensitive datasets:** tools for federated learning, differential privacy, and confidential computing.
- **Secure Processing Environments:** to ensure the safe and compliant training with sensitive data within the AI factories.

Data Pooling

- **Data discovery:** Identification of data across various sources and ecosystems, including datasets that are not yet fully discoverable, offering a data catalogue of the data that are or might become available through the Data Lab.
- **Data acquisition:** Support in obtaining data while considering diverse licensing models, contractual terms, and access and usage policies.
- **Data collection:** Establishing connections with national and European data repositories, data spaces, and other infrastructures that provide access to relevant datasets.
- **Metadata collection and management:** Access to the metadata and associated knowledge graph of the datasets available in the Data Lab. This might include semantics and ontologies.
- **Access to Public Data:** By storing relevant public dataset, Data Labs can offer immediate access large datasets.
- **Open Data Models:** Leveraging pre-trained open data models as valuable data sources for further AI model development.
- **Direct Link with common European Data Spaces.** Offering the possibility of accessing data available from data spaces through a direct link will allow AI developers to have access to large quantities of private data.
- **Creation of data pools:** Combining data from multiple sources into unified pools that can be made available beyond individual use cases, promoting reuse and scalability.

Regulatory clearance

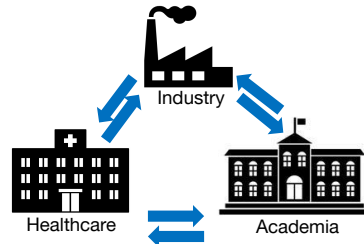
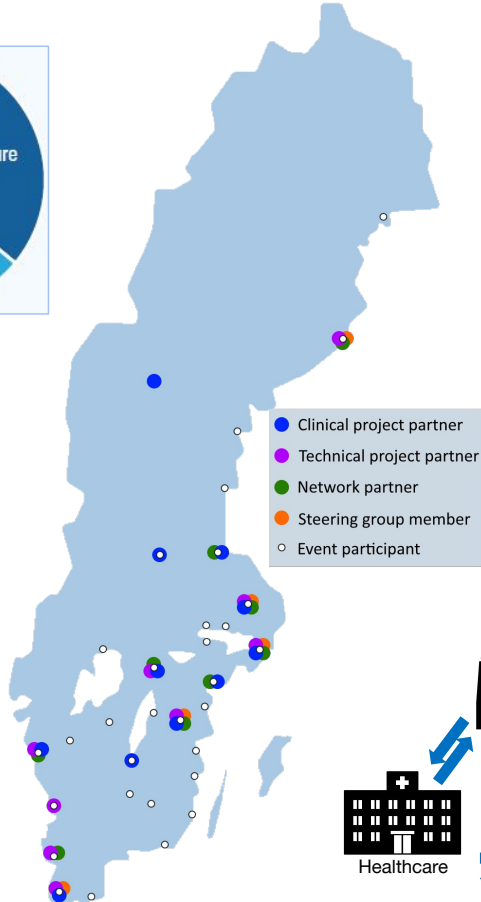
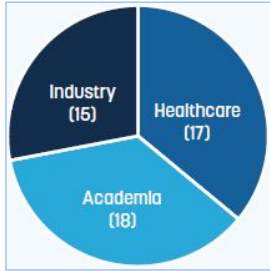
- **Guidance on regulatory compliance:** assistance in identifying the necessary steps to assure compliance with EU legislation (e.g. GDPR, trade secrets, copyright, and competition law, AI Act)
- **Sector-specific regulatory support.** Help in understanding and complying with domain-specific regulations (e.g. EHDS)

Training

- Data Labs should also provide training programs and self-learning resources to help AI developers effectively understand and utilise data analytics and AI tools.

2. Networking of Data Labs

Establishing a strong and coordinated network among the data labs is essential to maximize their collective impact. Specialisation enables deep expertise, but it also creates the need for collaboration to ensure broader coverage, knowledge exchange, cross-sector innovation.



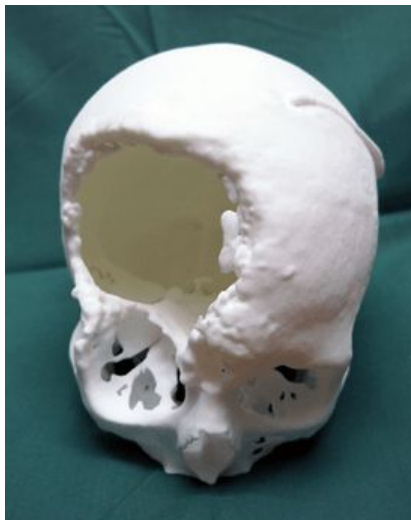
AIDA Community

Publicly funded collaboration arena for AI innovation in medical imaging diagnostics.

- ~50+ partners, 100+ projects since 2017
- 50+ workshops, 11 clinical AI courses
- Research & innovation projects
- Fellowships & Clinical evaluations
- Incubator for AI validation

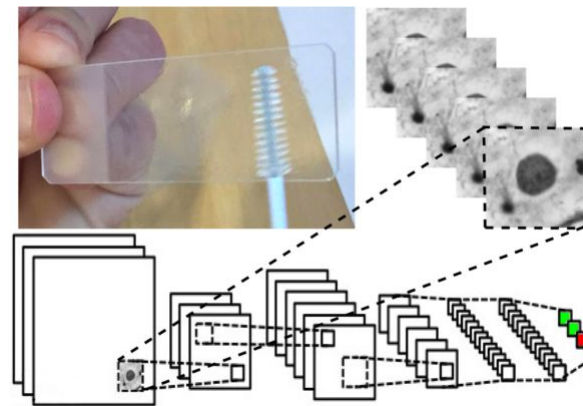
AIDA

Triple helix clinical innovation projects for patient benefit



Medviso
3D-printed replacement parts

Södersjukhuset
AI evaluation of intracranial hemorrhages



Uppsala University
Affordable oral cancer screening



AIDA Data Hub

E-infrastructure for research and clinical innovation in data driven precision health.

Data services

- Access high quality datasets
- FAIR sharing of DOI citable datasets
- Extract and enrich clinical data for research

Data Science Platform for Sensitive data

- Secure long term primary storage & compute.
- Advanced usage patterns:
collaborate, annotate, federate, train AI, ...

Support

- Data sharing, ethics, legal, and policy
- AI development & System design





AIDA Data Hub

Data Science Platform

A home for your research and innovation in data driven precision health.

Secure data science platform co-located with national/European flagship compute systems.

Supporting advanced data usage patterns: long term primary storage, collaborate, annotate, share, federate, train AI...

Customers make security decisions as appropriate; outgoing connections to home institution servers, collaborators...

User fees for sustainable operations and development. Discounts to incentivize data sharing and maximize high impact research.





Data Out

Metrics:

- Countries: 47
- External sharing events: 328



	Regional lymph node metastasis in colon adenocarcinoma Keywords: Pathology, Whole slide imaging, Annotated, Lymph nodes, Cancer, Colon, Adenocarcinoma. doi:10.23698/aida/Inco	2019
	Regional lymph node metastasis in colon adenocarcinoma, second collection series Keywords: Pathology, Annotated, Whole slide imaging, Lymph nodes, Cancer, Colon, Adenocarcinoma. doi:10.23698/aida/Inco2	2020
	Mammograms from cancer screening Keywords: Radiology, Breast, Cancer, Mammography, MMG, MG, MAMMO, DM. doi:10.23698/aida/mbtst-dm	2019
	Multi-echo gradient echo (MEGRE) Keywords: Gold fiducial marker, MRI only, Radiotherapy, MRI radiation therapy, Prostate, Cancer, Radiology, Annotated. doi:10.23698/aida/megre	2020
	Coregistered H&E- and VEGF-stained histology slides with diffusion tensor imaging data at 200 μm resolution in meningioma tumors Keywords: Radiology, Microstructural features, Mean diffusivity, Fractional anisotropy, cCell density, Cellularity, Meningioma, Coregistration, Hematoxylin & eosin, VEGF. doi:10.23698/aida/micromen	2023
	SCAPIS body data from site 1 Keywords: Radiology, Computed tomography, Body, Scapis. doi:10.23698/aida/scapis-ct-site-1-body	2024
	SCAPIS CASC data from Site 1 Keywords: Radiology, Computed tomography, Heart, Scapis. doi:10.23698/aida/scapis-ct-site-1-casc	2024
	SCAPIS CCTA data from site 1 Keywords: Radiology, Computed tomography, Heart, Scapis. doi:10.23698/aida/scapis-ct-site-1-ccta	2024

 **Sök**

Per sida: 10 ▾

Publicerad: 2019-09-12



Publicerad: 2019-08-13



Tillgång till data är begränsad







Illustration from Terminator Genisys (2015).

Statistical Learning

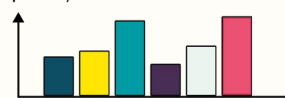
World



Some data generating process where we can only observe events (e.g. it's now red), not the how the events are produced.

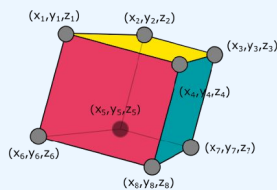
Observed event

Frequency

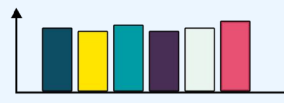


Statistics of observed events

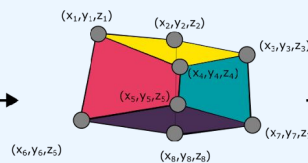
Model



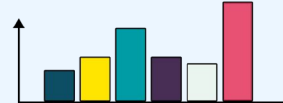
Probabilistic *model* of the data generating process with "parameters" we can adjust



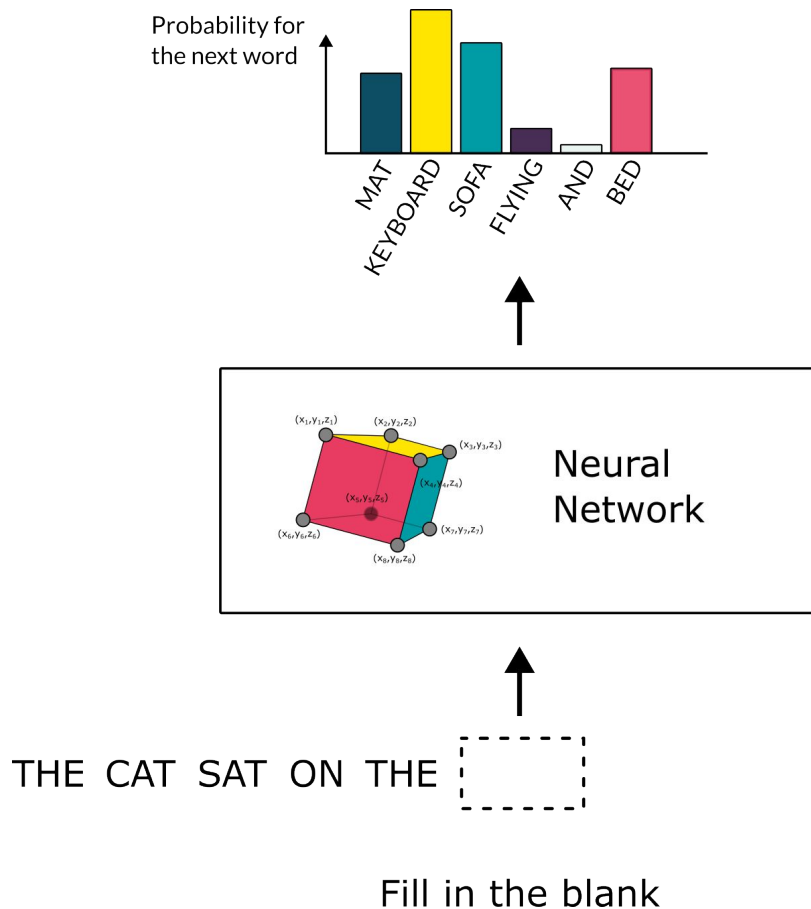
Statistics of models predicted events



Adjusted model which generates data similar to the training data



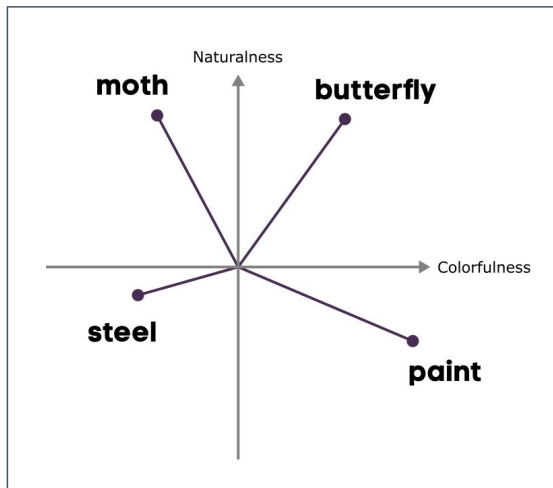
Statistics of adjusted models predicted events



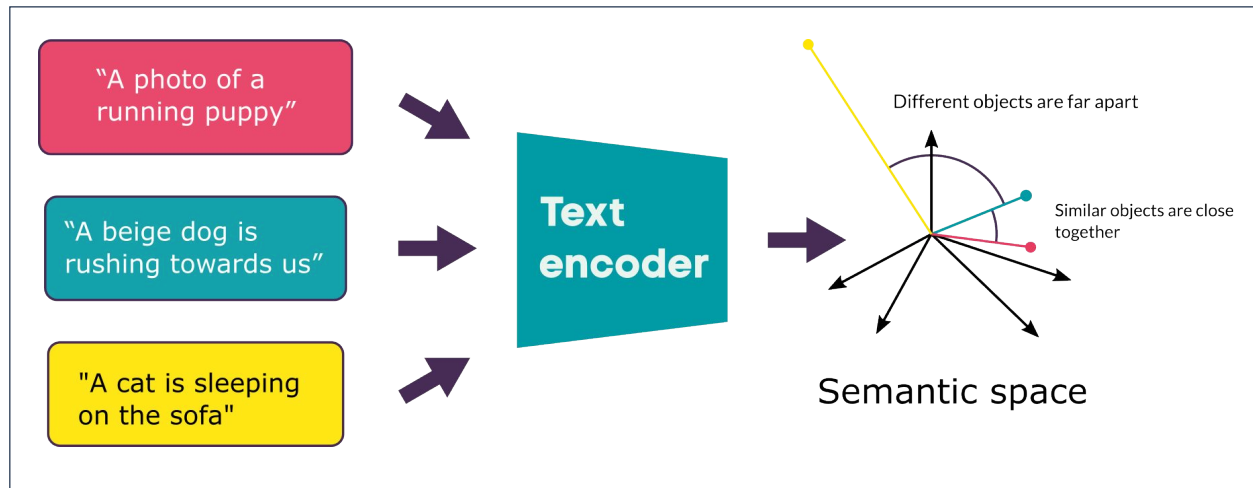
Langue modelling fine tuning

- Existing model (XLM-RoBERTa) is fine-tuned on de-identified clinical notes
- XLM-RoBERTa is a multilingual model trained on 2.5TB or text from 100 languages.
- Trained using “Masked” Language Modelling, model fills in masked words to learn word vectors based on surrounding context
- Clinical notes are divided into context windows (512 tokens, ~400 words) the model can handle
- Fine-tuned model is used to create “contextual” word vectors for all words in the training data set, marking them as being in the implant list or not

Semantic vector search



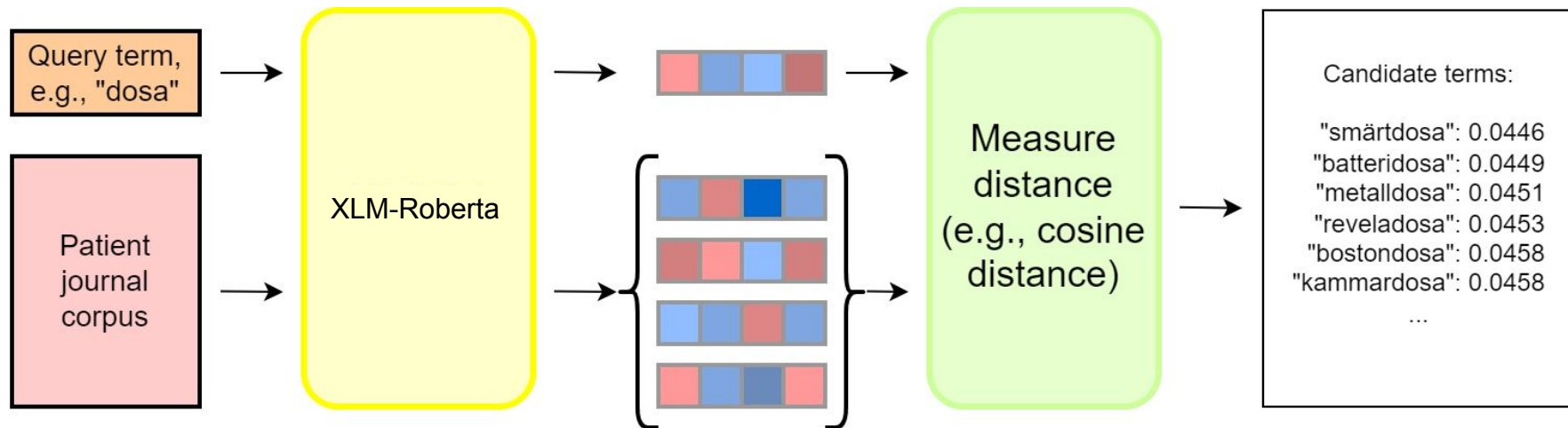
Word vectors – each word is encoded as a vector in a space, where **different directions** in space encode **different semantics**



Similar objects represented as vectors can be found by looking at how close they are in the semantic space

Semantic search with (BERT-style) LLM

- We can find new glossary terms from the clinical notes by searching for words which are similar to glossary words or are used in a similar context.
- Similarity can be assessed by measuring the distance between the embeddings of different terms.
- XLM-Roberta will be used for representations



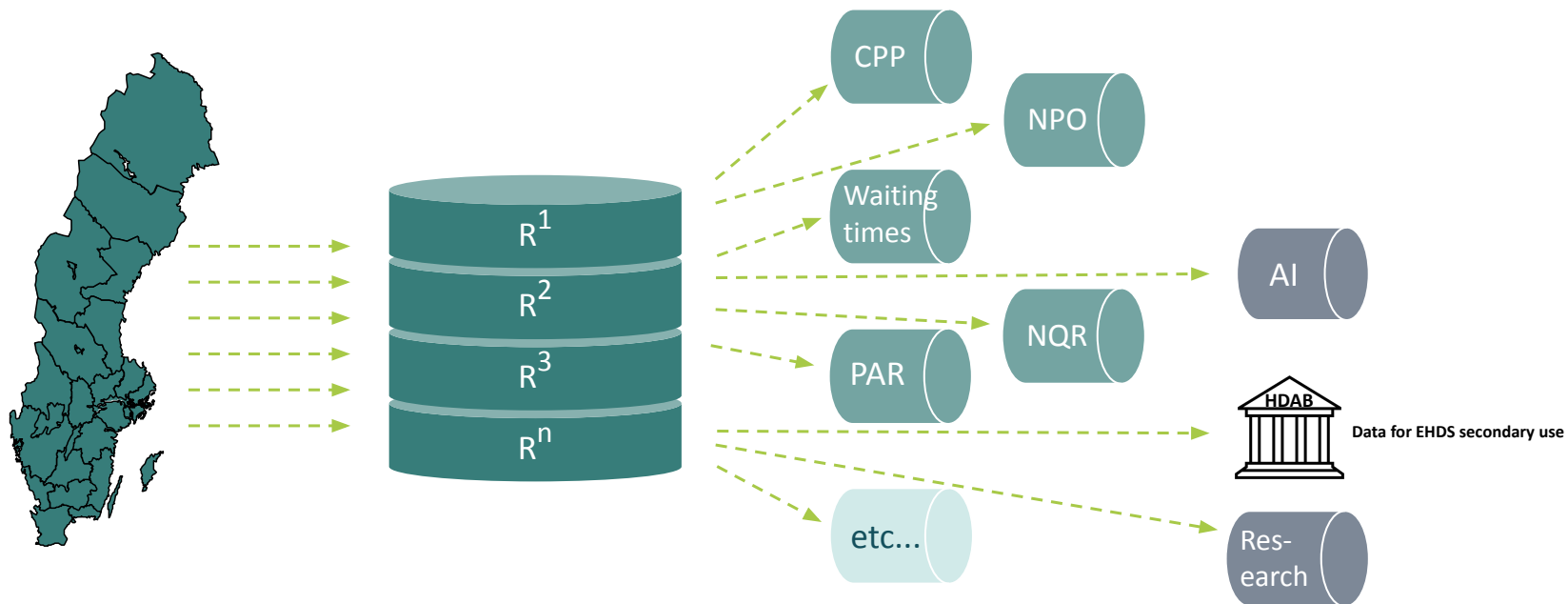


ASHA - Använda Standardiserade Hälsodata som Accelerator

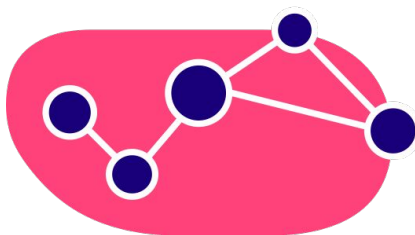
VINNOVA Systems demonstrator for healthcare transition to open standards for health data.

AIDA Data Hub provides secure environments for primary and secondary use.

Vision: Common national healthcare data infrastructure



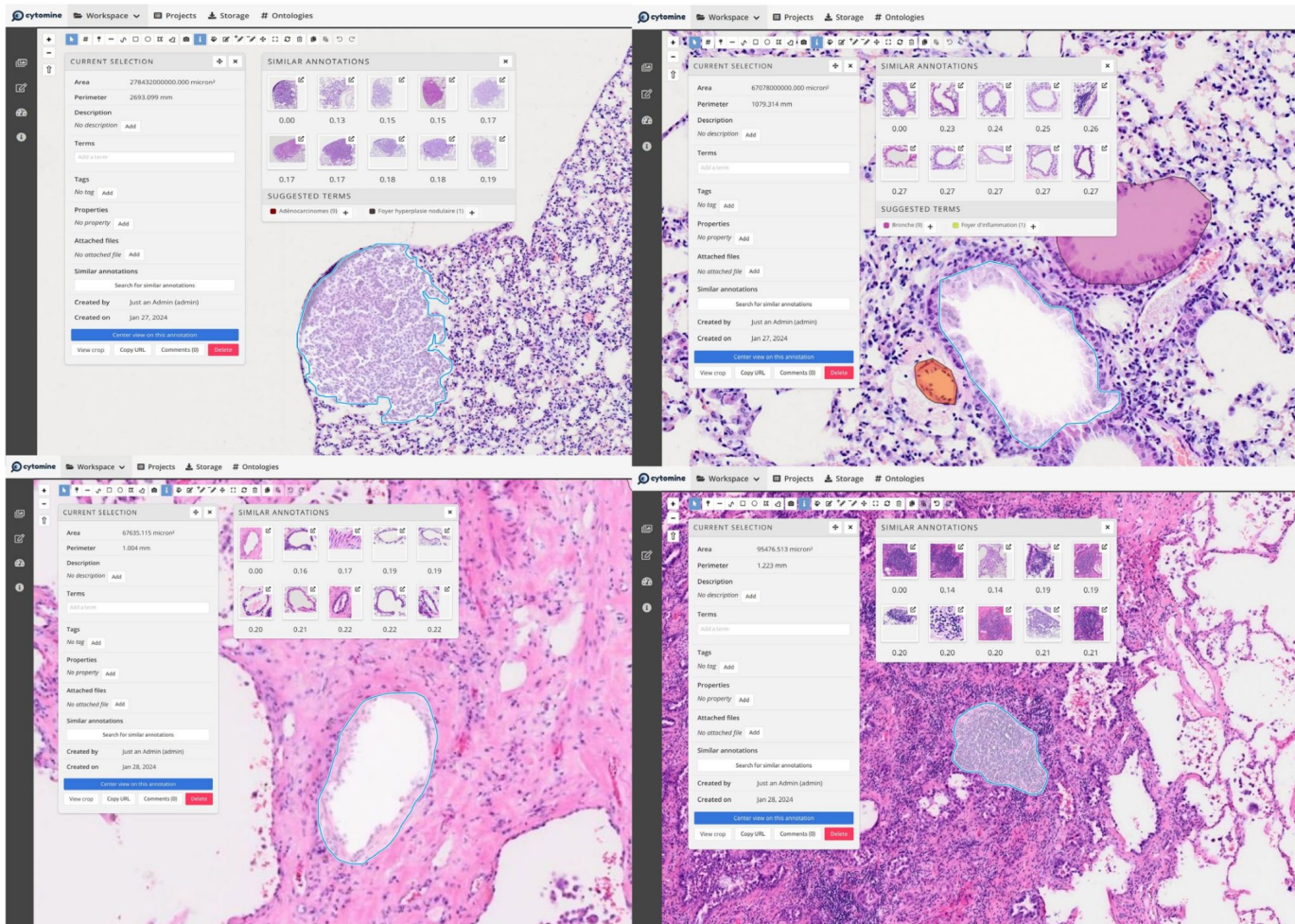
bigpicture



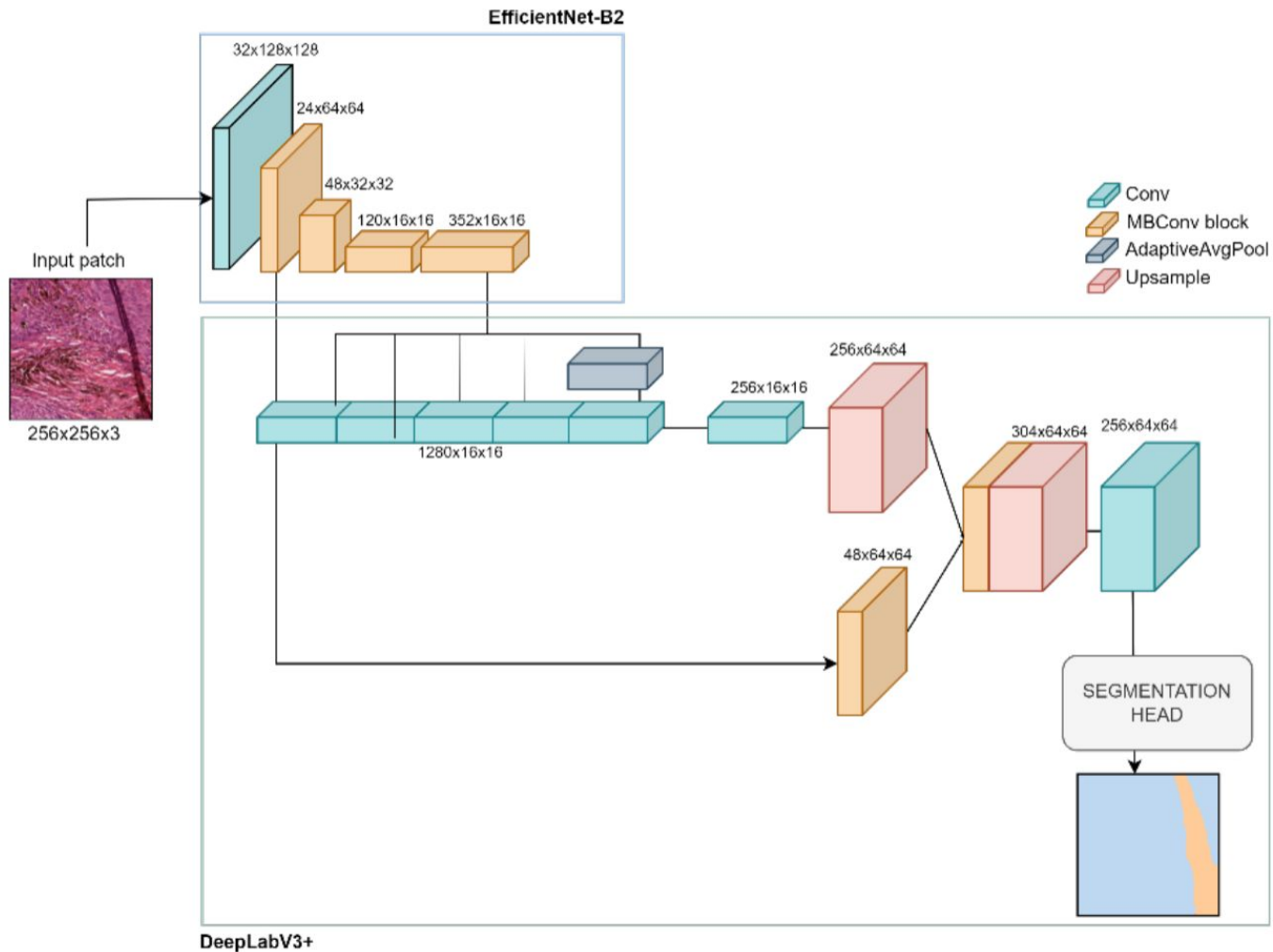
Bigpicture **Petabyte platform for European digital pathology AI**

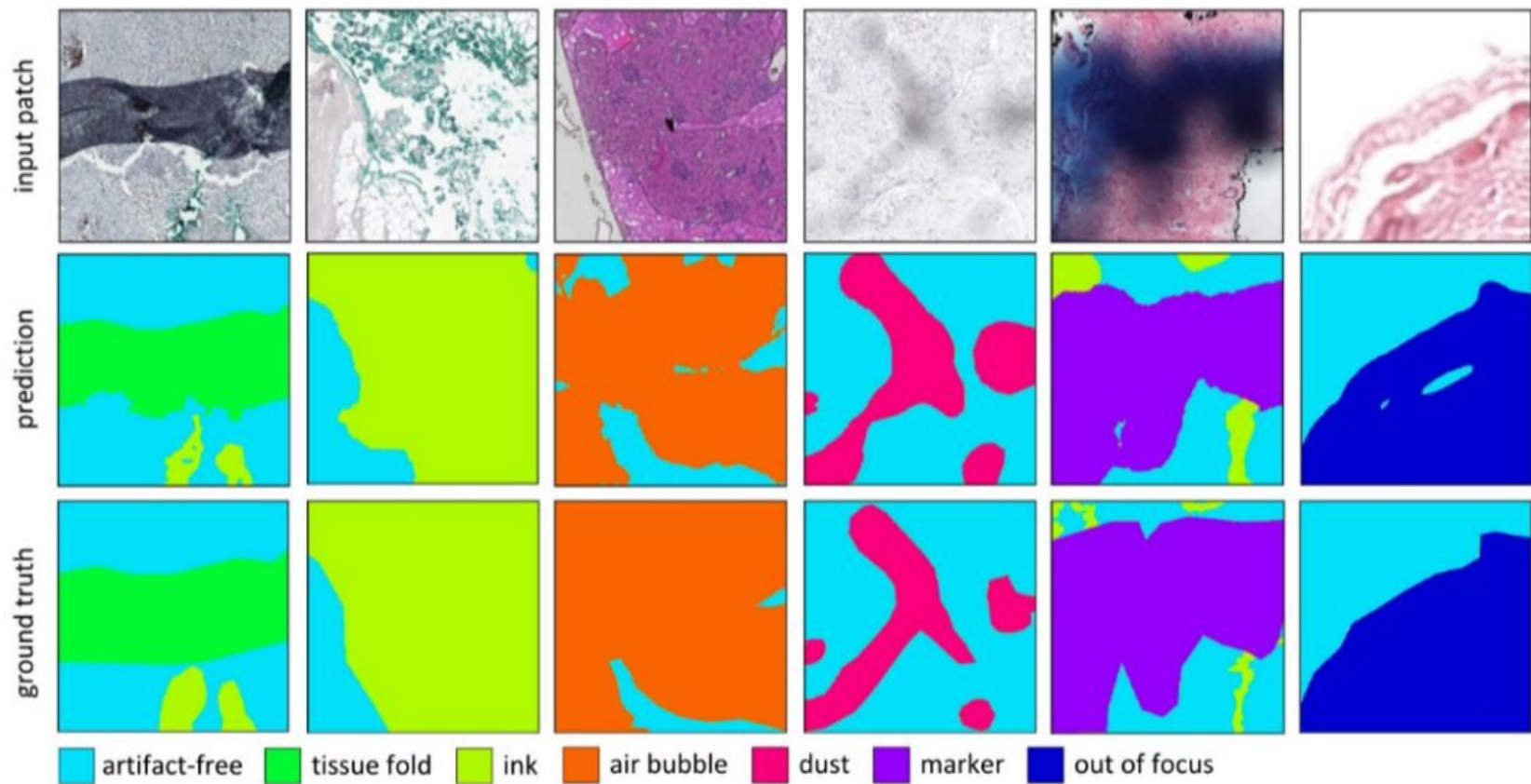
AIDA Data Hub leading repository infrastructure development, based on FedEGA / GDI technologies in collaboration with sensitive data teams at ELIXIR-SE and -FI.

AIDA Data Hub Data Science platform to host the first Bigpicture Federated node.

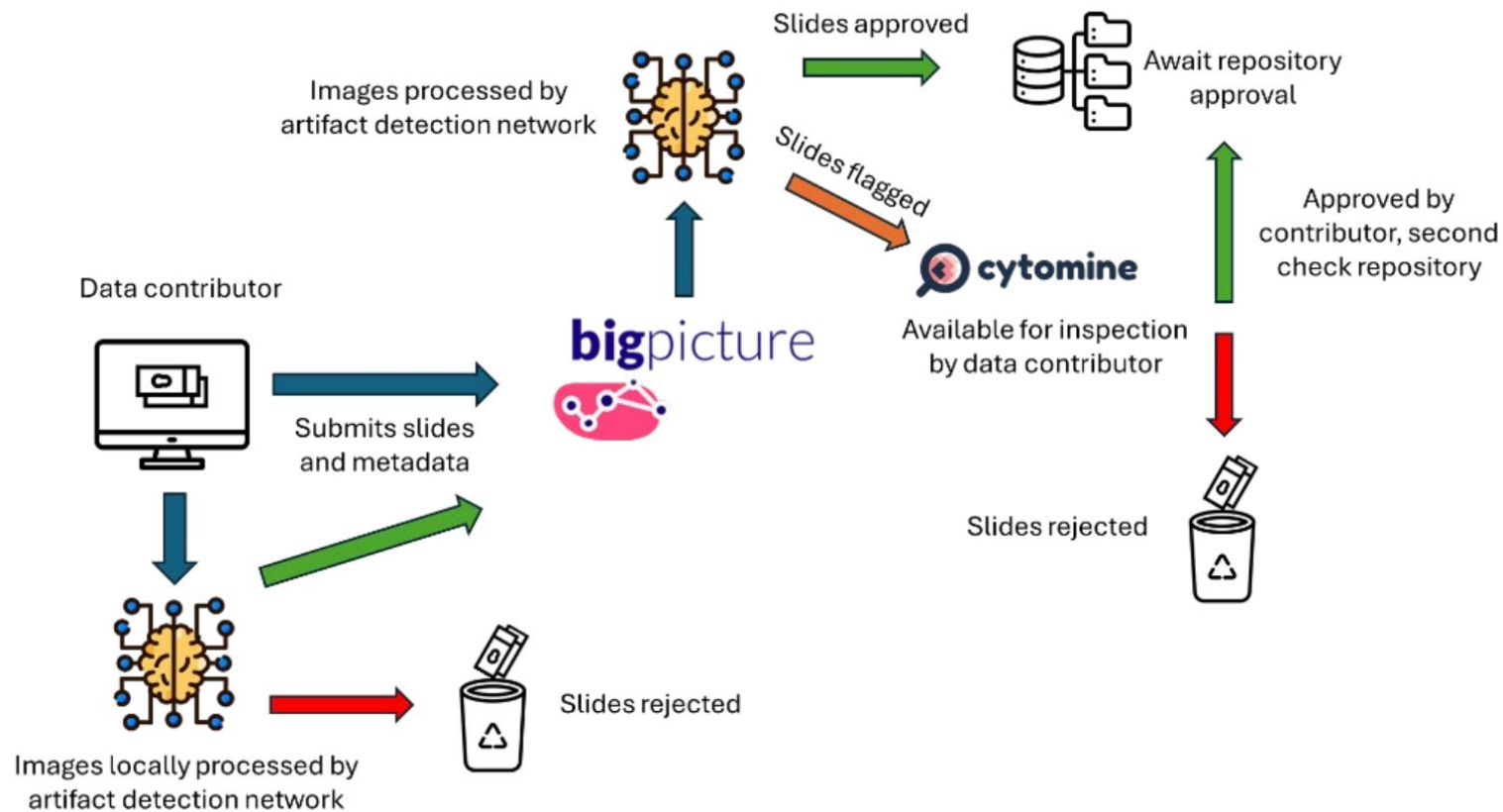


Graphic from Bigpicture - European repository for AI in pathology.





Platform integration





MIMER

Storskaligt användande av AI

med fokus på tillgängliggörande av metadata
som du inte visste att du hade

Joel Hedlund, NAISS Data Science Director

NAISS



RI
SE



MIMER

MIMER

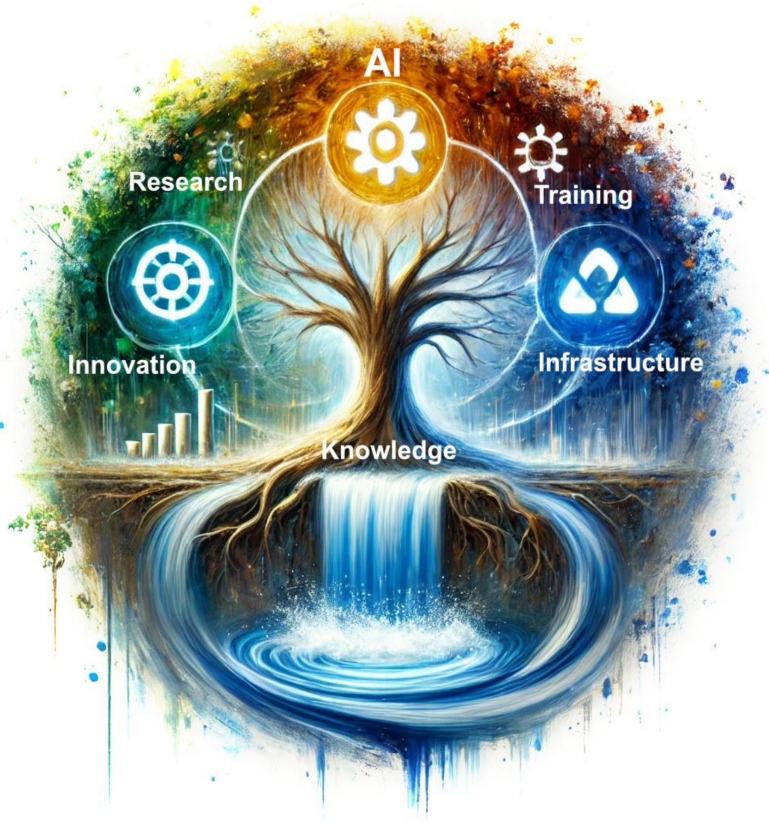
AI FACTORY

NAISS



**RI
SE**

Extra slides in case of questions...



2025-2027+

I know where Othin's eye is hidden, Deep in the wide-famed well of Mimir;

Mead from the pledge of Othin each morn Does Mimir drink: would you know yet more?

--Snorri Sturlasson, 13 century



A pure AI system

Cloud native environment.

Kubernetes, IaaS, SaaS, instant resources.

Jupyter, web interfaces.

AI on tap, for standardized workloads.

More intuitive to non-HPC users?

and much more...

RESEARCH

Open Access



European health regulations reduce registry-based research

Oscar Brück^{1*}, Enni Sanmark², Ville Ponkilainen³, Alexander Bützow⁴, Aleksi Reito³, Joonas H. Kauppila^{5,6} and Ilari Kuitunen⁷

Abstract

Background The European Health Data Space (EHDS) regulation has been proposed to harmonize health data processing. Given its parallels with the Act on Secondary Use of Health and Social Data (Secondary Use Act) implemented in Finland in 2020, this study examines the consequences of heightened privacy constraints on registry-based medical research.

Methods We collected study permit counts approved by university hospitals in Finland in 2014–2023 and the data authority Findata in 2020–2023. The changes in the study permit counts were analysed before and after the implementation of the General Data Protection Regulation (GDPR) and the Secondary Use Act. By fitting a linear regression model, we estimated the deficit in study counts following the Secondary Use Act.

Results Between 2020 and 2023, a median of 5.5% fewer data permits were approved annually by Finnish university hospitals. On the basis of linear regression modelling, we estimated a reduction of 46.9% in new data permits nationally in 2023 compared with the expected count. Similar changes were neither observed after the implementation of the GDPR nor in permit counts of other medical research types, confirming that the deficit was caused by the Secondary Use Act.

Conclusions This study highlights concerns related to data privacy laws for registry-based medical research and future patient care.

Key Points

- Given its parallels with the European Health Data Space regulation, we modelled its possible consequences for registry-based medical research using data from Finland, where the Act on the Secondary Use of Health and Social Data (Secondary Use Act) has been implemented since 2020.
- Adjusted for the historical trend of increasing registry-based research conducted in Finnish university hospitals, the data permit count was almost 50% lower than expected in 2023.
- Similar changes were not observed post-GDPR or in other medical research types.
- This study demonstrates how increased data privacy regulations might reduce medical research, innovations and advances in future patient care.

*Correspondence:
Oscar Brück
oscarbruck@hus.fi
Full list of author information is available at the end of the article

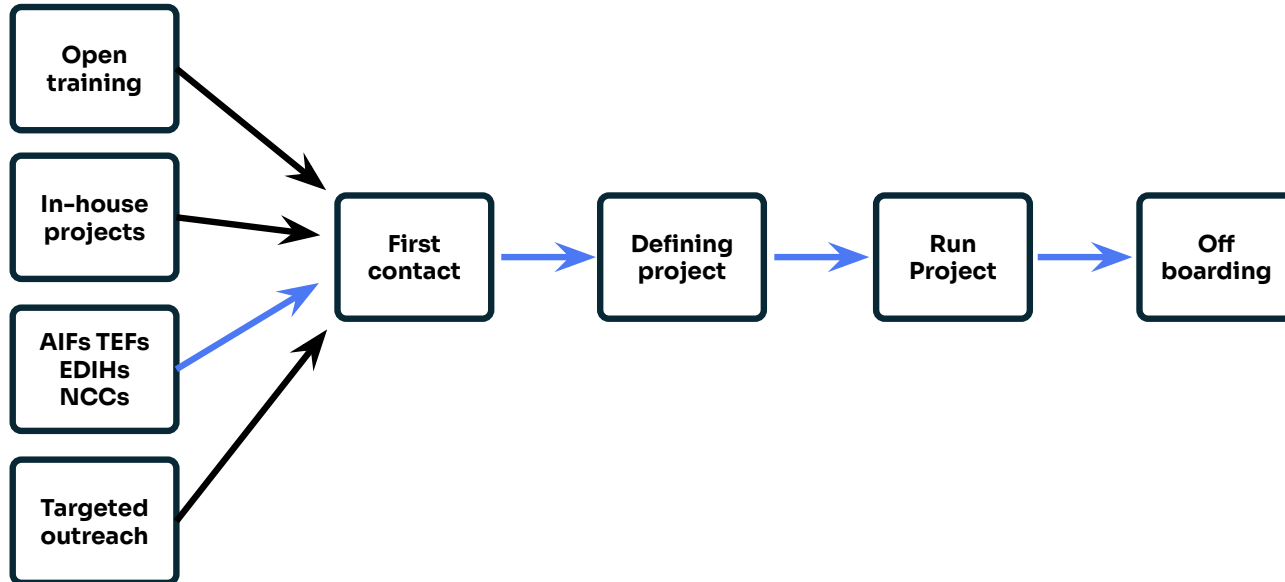


© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.



Customer journey

Entry points





What do I get as a researcher?

Assistance in getting free access to compute and storage resources.

Support and training:

- On-boarding.
- Data management.
- Model architecture and optimisation.
- Integrated AI and domain expertise.
- Networking and co-working hubs
- Thematic workshops



Hardware parameters

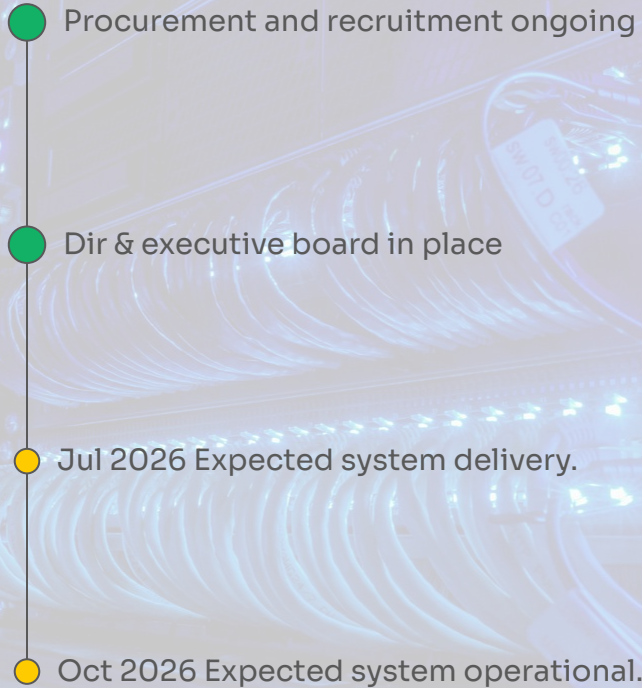
Procurement discussions with EuroHPC underway.

Cloud first.

Sensitive data ready.

GPU for AI training and inference.

Massive object storage, fast NVMe cache.

A vertical timeline is positioned on the left side of the slide. It features a central vertical line with four circular markers. The top two markers are green, and the bottom two are yellow. To the right of each marker is a text label. The background of the slide is a blurred image of server racks with glowing blue lights.

● Procurement and recruitment ongoing

● Dir & executive board in place

● Jul 2026 Expected system delivery.

● Oct 2026 Expected system operational.

Timelines

Director and executive board in place.

Recruiting 52 FTE AI experts, kick-started by ENCCS, NAISS, and affiliated parties.

Procurement discussions with JU underway.

Expected hardware delivery in Jul 2026.

Expected operational in Oct 2026.



AIDA Data Hub Data Science Platform

A home for your research.

534547657568
675756756756
7867876889
7878678789789
87798797
7867886976
78979878978

Arrhenius & Mimer

EuroHPC

NAISS



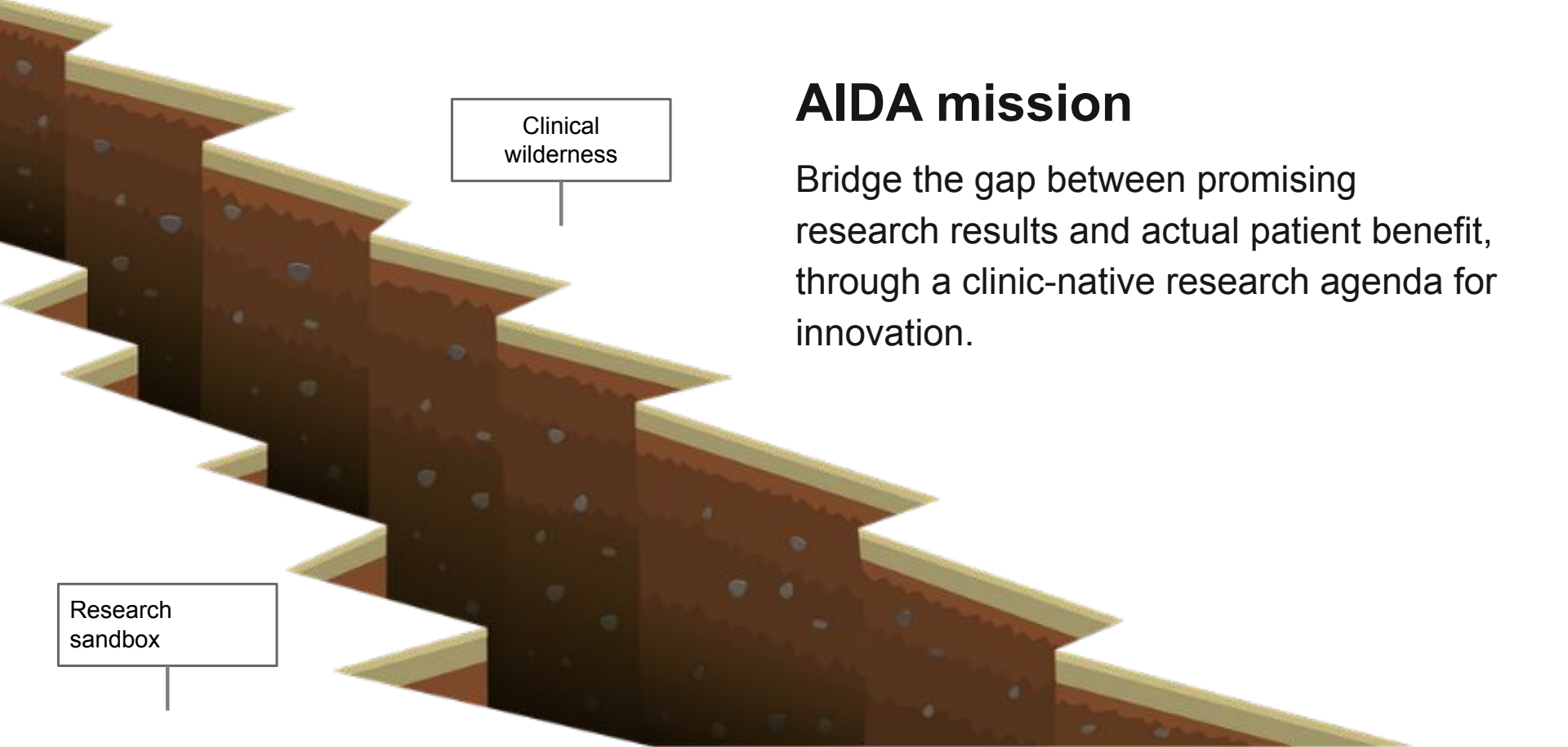
AIDA & AIDA Data Hub

AIDA Community - medtech4health.se/aida

National collaboration arena for AI research and innovation in medical imaging diagnostics.

AIDA Data Hub - datahub.aida.scilifelab.se

E-infrastructure set up to support AIDA.



Clinical
wilderness

Research
sandbox

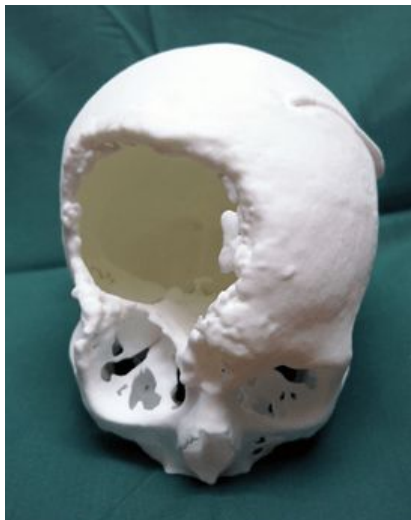
AIDA mission

Bridge the gap between promising research results and actual patient benefit, through a clinic-native research agenda for innovation.

23423435345464
5446565464656646
657656567
786768
67866876876
786768678
786767

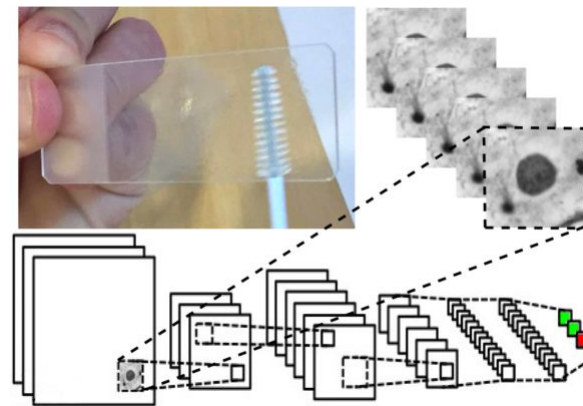
AIDA

Triple helix clinical innovation projects for patient benefit

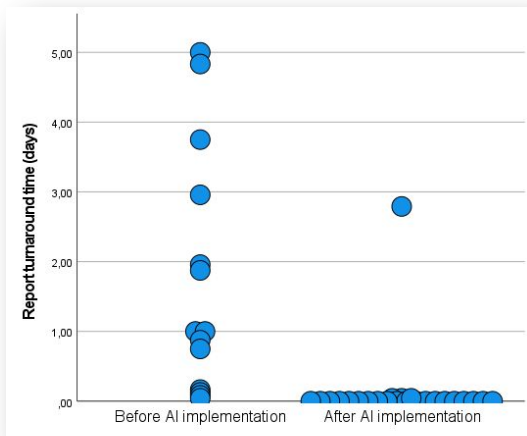


Medviso
3D-printed replacement parts

Södersjukhuset
AI evaluation of intracranial hemorrhages



Uppsala University
Affordable oral cancer screening



Success story: Region Halland

- Participated in AI course
- Participated in AI showcase event
- Interest in pulmonary embolism tool
- Started clinical evaluation and implementation
- 2022: Patient benefit achieved





EUCAIM

Federated infrastructure for cancer imaging data

AIDA Data Hub contributing data collaboration environments for use in EUCAIM with cancer imaging data based on Bigpicture Federated node technologies.

Collaboration with sensitive data teams at the NBIS Systems Development unit.



Data Sharing Policy

[AIDA Data Sharing Policy](#)

Comprehensive resource describing best practices in handling and sharing medical imaging data for research in Sweden and similar countries.

Concrete guidelines and examples, with references to original sources in law.

Developed multi-professionally with AIDA network and stakeholders.

Key insights have been published in [Nature Scientific Data](#).





Using Clinical Imaging Data for Research

Common practice in Sweden and similar countries, 1-paragraph summary:

The common practice is that **caregivers disclose** data to **research institutions** for specific activities described in approved **ethical review applications**, to be carried out under appropriate technical and organizational **protective measures** and supervised by a named **competent researcher**. The research institution is then **data controller** and **copyright holder** for the disclosed data, and is responsible for ensuring that data is processed and shared only as described in the approved ethical review application, with **data processing agreements**, **pseudonymization**, **anonymization** and **licensing** as tools, and with an obligation to store relevant data for 10 years after last use for purposes of **research validation**.



MIMER

AI FACTORY



MIMER

NAISS



RI
SE