gaia-x

Health Data Space Event

4 APRIL 2022

Chapter 2a: Breakout Use cases

gaia-x

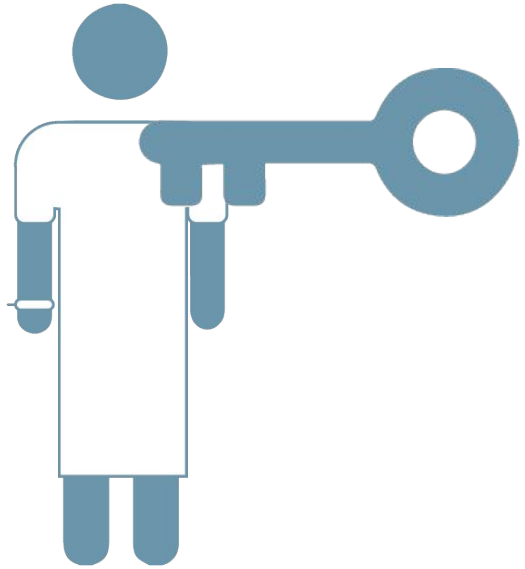
#6 Enabler trusted AI

Welcome & opening

- **Prof. Ulf Nehrbass, CEO,** Luxembourg Institute of Health

AI in healthcare – the transborder Clinnova initiative

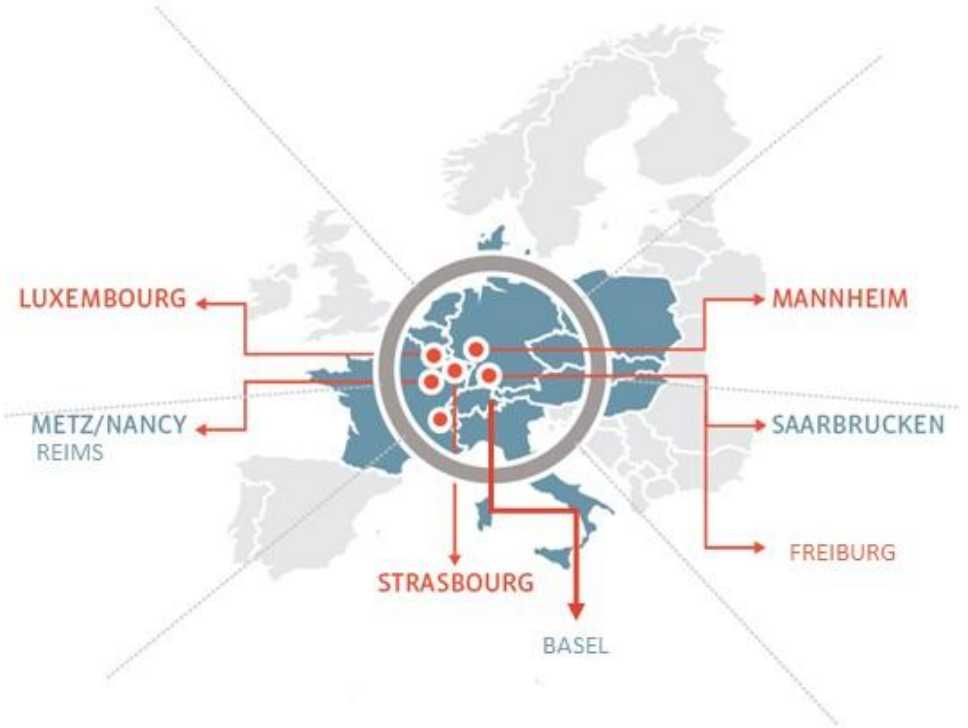
Cross-border digital health hub



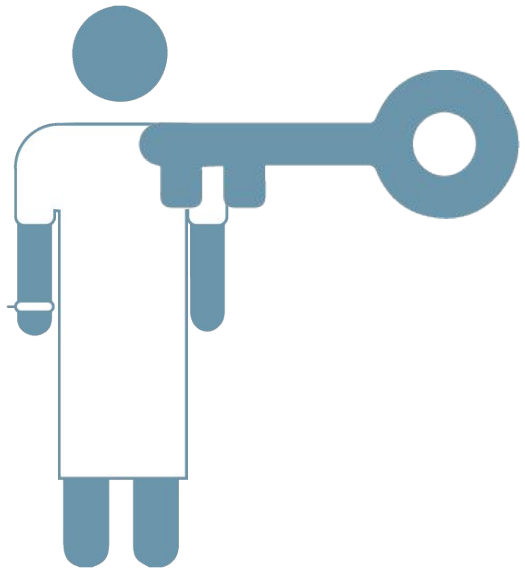
CLINNOVA'S STRATEGIC GOAL:

- Unlock the potential of data science and artificial intelligence (AI) in health care

Clinnova partners



Cross-border digital health hub



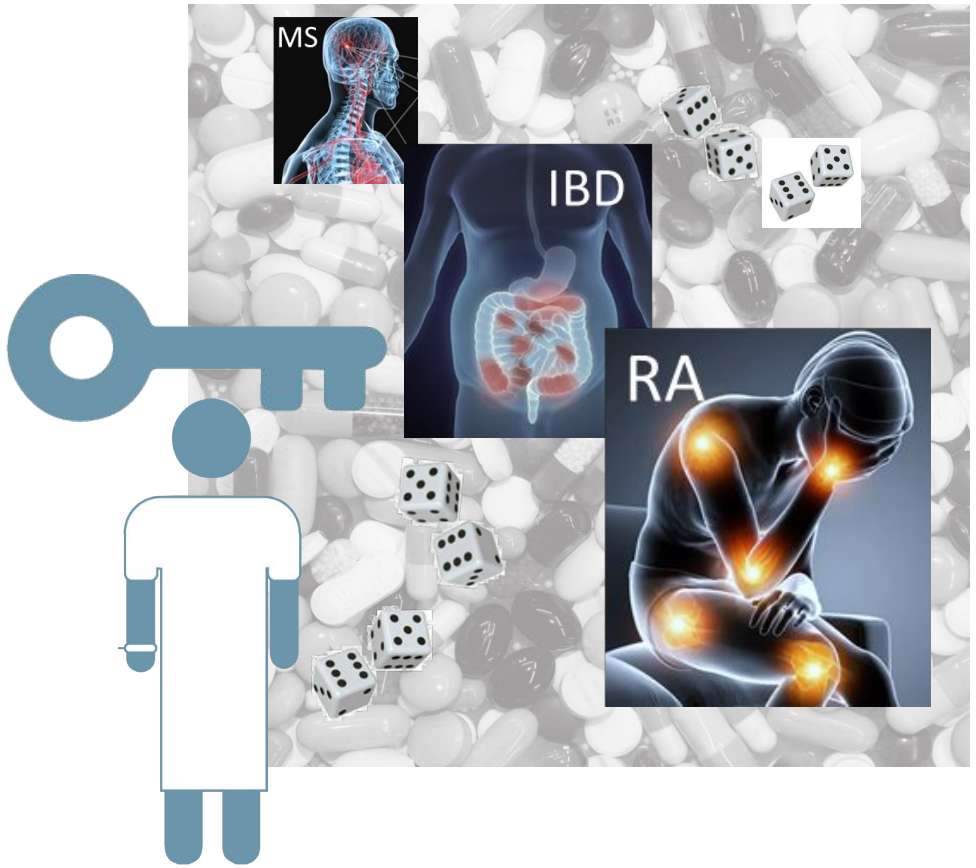
CLINNOVA CONSIDERS AI-INNOVATION IN HEALTHCARE AN OPERATIONAL AND ORGANIZATIONAL CHALLENGE

- The key innovation driver is not the AI or ML algorithm
- It rather lies in a data-enabling environment that produces standardized, quality-controlled data around relevant use cases

Cross-border digital health hub

CLINNOVA'S PROSPECTIVE STUDIES START WITH 3 MEDICAL USE CASES

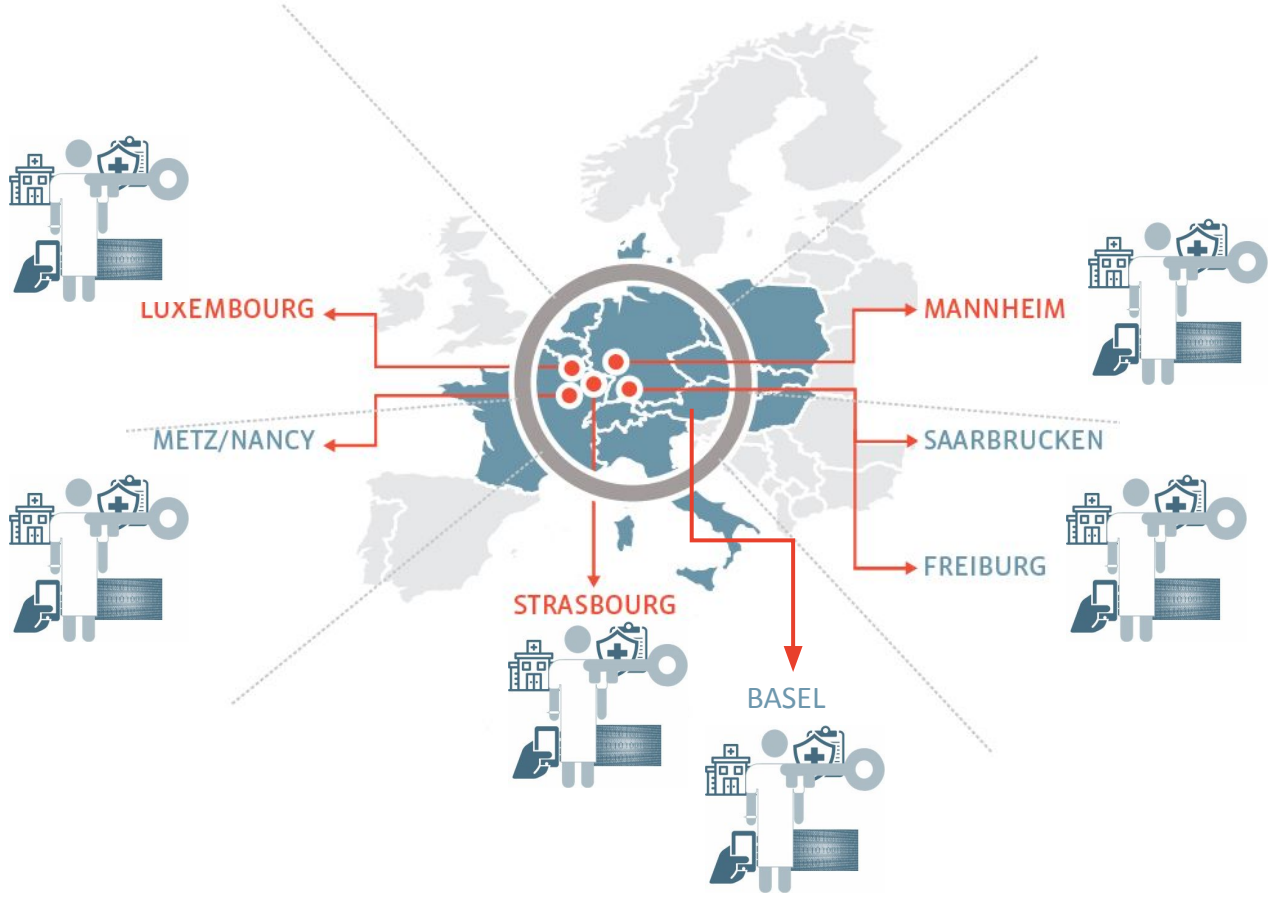
- RA, IBD and MS: 30Mio patients, €55Bn/ year market
- Standard of care drugs are abound, but it is not known which drug benefits what patient
- Clinnova's stratification approach will assign the right drug to the right patient



Clinnova trans-border dimension

FACILITATING TRANS-BORDER DATA FLOWS AND COLLABORATION

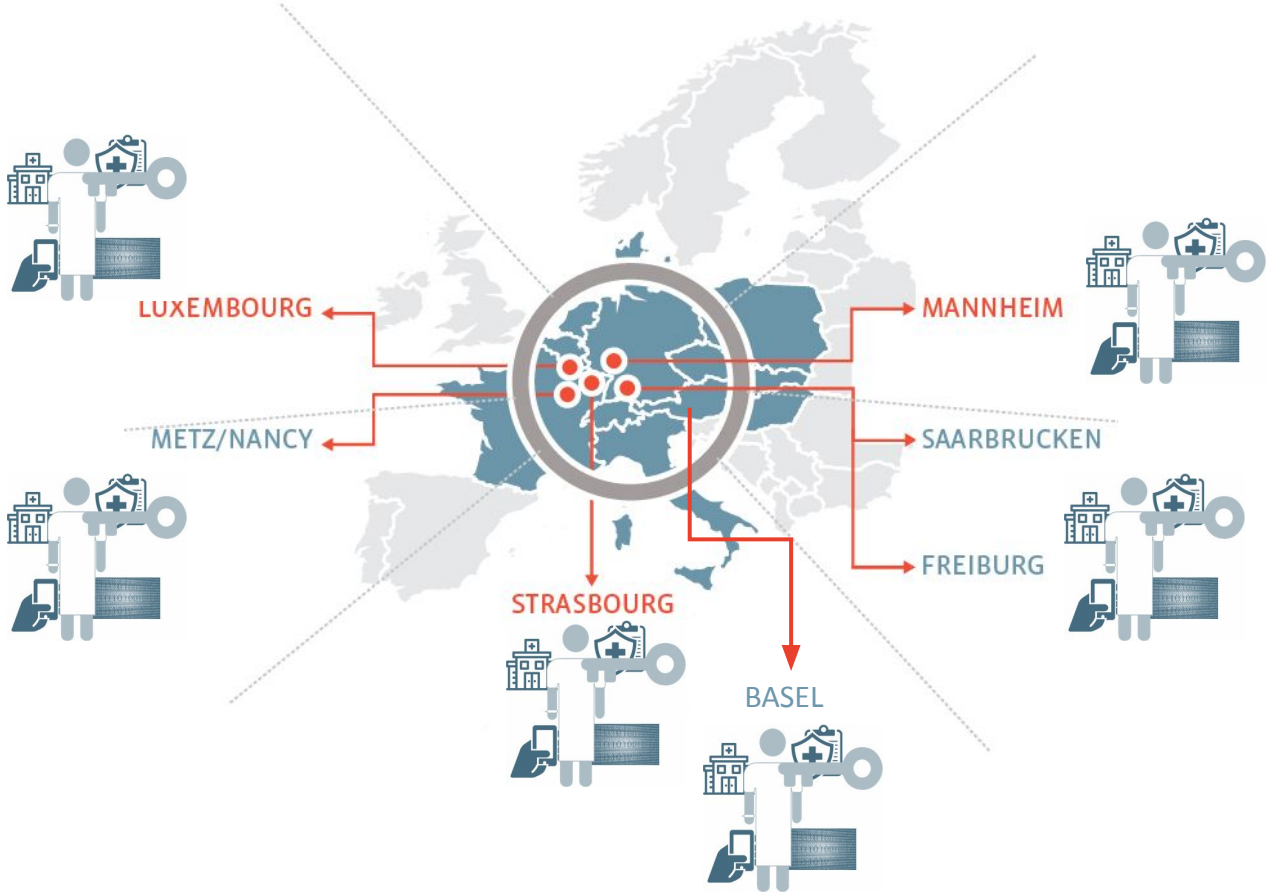
- Linking up data integration platforms between clinical and research centers across borders
- Federated computing: data remain in local data integration centres and analyses are performed remotely
- Links health data context with neighbours to assure interoperability



Clinnova trans-border dimension

CRITICAL QUESTIONS ON TRUSTED AI

- How can we ensure trustworthiness of AI algorithms?
- How can we achieve self-explainability of algorithms?
- How can we create a data ecosystem in which data do not need to be transferred / exchanged?



Strategic discussion on Trusted AI

- **Prof. Philipp Slusallek,** Scientific Director, German Research Center for Artifici Intelligence (DFKI)
- **Dr Christian Schorr,** Senior Researcher, DFKI

Trusted AI

gaia-x

– How advanced anonymization can improve federated architectures

- **Tuomo Pentikäinen,** CEO, VEIL.AI

TRUSTED AI – HOW ADVANCED ANONYMIZATION CAN IMPROVE FEDERATED ARCHITECTURES

GDPR-free row level health data through *new next generation anonymization technology*

Tuomo Pentikäinen, CEO

VEIL.AI

April 4th, 2021

VEIL.AI

SHORT COMPANY INTRODUCTION:

VEIL.AI

ENABLING USE OF SENSITIVE DATA

Pseudonymized data

Anonymized data

Synthetic data

OUR BACKGROUND

VEIL.AI HAS DEEP ROOTS IN SENSITIVE HEALTH DATA

FIMM
Institute for Molecular Medicine Finland
Nordic EMBL Partnership for Molecular Medicine

- FIMM – Institute for Molecular Medicine Finland – with 230 employees representing tens of nationalities, is ***focusing on human genomics and precision medicine***

  (Under the umbrella of the Helsinki Institute of Life Science at the University of Helsinki)

Key Data Scientists

- Our key technology experts have an ***experience of 20+ years*** related to demanding sensitive data management projects, such as:

MultipleMS FINNGEN GeneRISK DIGITAL HEALTH REVOLUTION HUB Biobank

VEIL.AI
ENABLING USE OF SENSITIVE DATA

- VEIL.AI a spin-out from FIMM, a health technology company located at Helsinki Meilahti hospital campus
- Founded in 2019
- Main customer groups: Global Pharma companies, University Hospitals, Health Regulatory bodies, MedTech companies

OUR SOLUTION:

VEIL.AI HAS DEVELOPED A GAME-CHANGING AI-BASED TECHNOLOGY THAT UNLEASHES THE POWER OF DATA

VEIL.AI Anonymization Engine

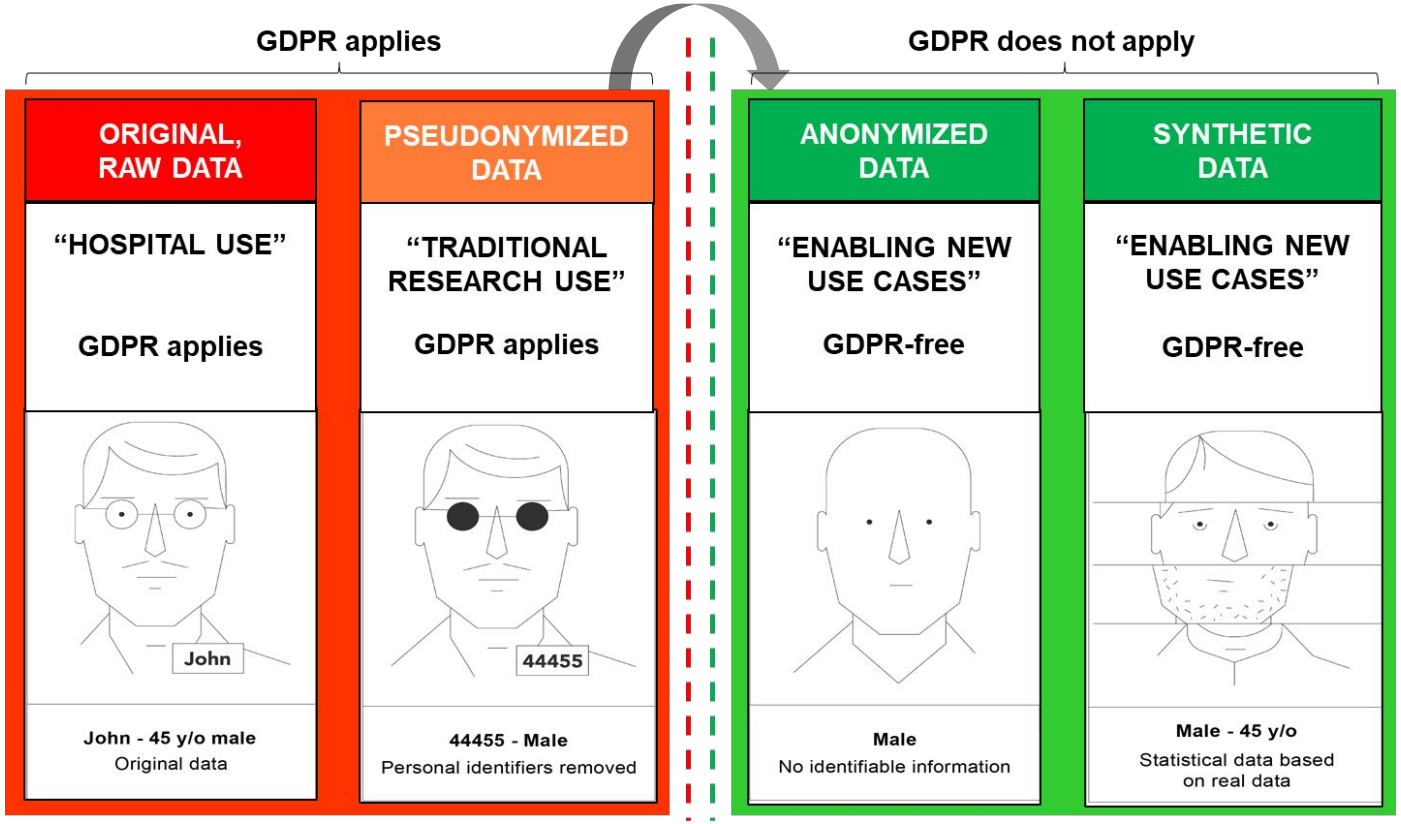
software

- The scalable *VEIL.AI Anonymization Engine* creates ***extremely high-quality*** ***row level anonymized and synthetic data***, providing great new opportunities for
  - Hospitals
  - Health data hubs / Regulatory bodies
  - Pharma and Diagnostics companies
  - All organizations utilizing sensitive data

WHY TO USE ADVANCED ANONYMIZED DATA?

- **One of the big advantages of anonymized data is that according to GDPR it is not considered as personal data (= it is "GDPR-free"). Therefore access and utility of health data improves significantly especially in transborder data collaborations.**

- **By utilizing also advanced anonymized data the organizations can improve the quality and quantity of the data they need (eg. clinical data).**



- We can see a shift in demand from pseudonymized data to advanced anonymized data
  - Due to "GDPR-freedom"

GDPR Recital 26

…

… "The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable.

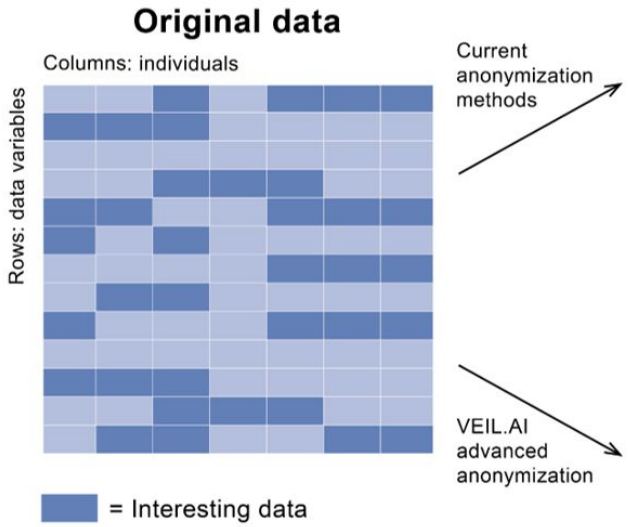
This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes."

VEIL.AI

COMPARISON

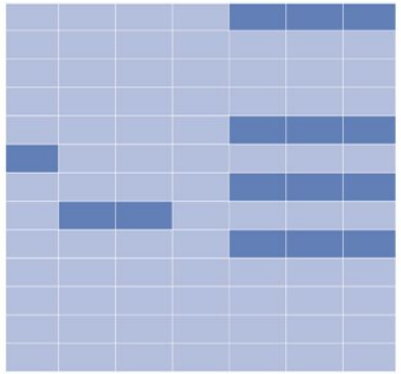
NEW ANONYMIZATION TECHNOLOGY vs. OLD METHODS

VEIL.AI's unique technology enables advanced anonymized datasets that can be used to draw the same conclusions as from the original data



Current anonymization methods

Old state-of-the-art methods



- Transformation loses interesting data
- Much of the interesting data has disappeared
- It is difficult to draw the same conclusions as from the original data

VEIL.AI advanced anonymization

VEIL.AI method

- VEIL.AI finds the best anonymization solution
- The interesting data is still available
- We can draw the same conclusions as from the original data

AI based, next generation anonymization technology

VEIL.AI

FEDERATED LEARNING ARCHITECTURE CAN BE A GOOD APPROACH

TRANSBORDER DATA COLLABORATION INCLUDE MAJOR DATA PRIVACY AND DATA SECURITY RISKS.

SOME SUGGEST THAT FEDERATED LEARNING SOLVES ALL THESE PROBLEMS.

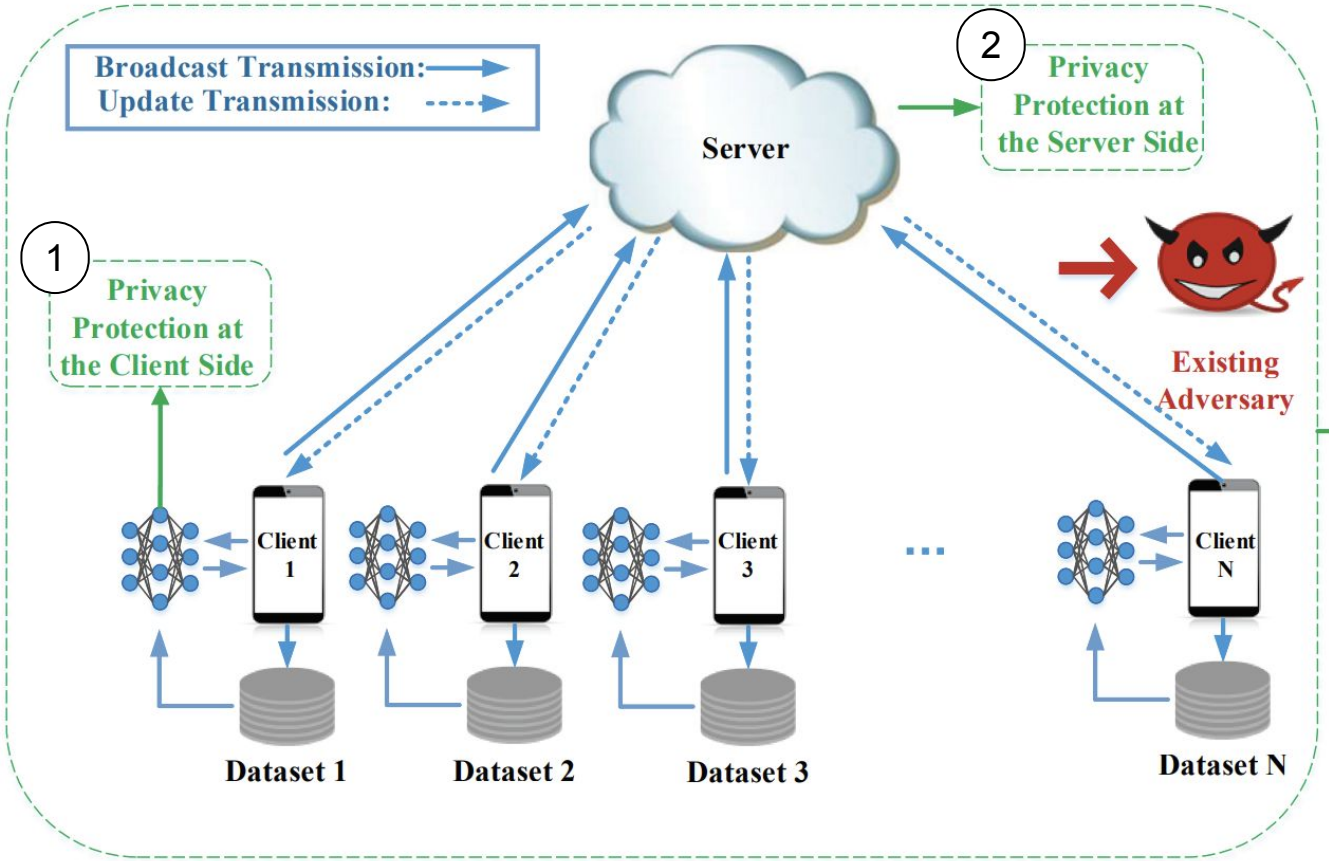
UNFORTUNATELY THIS IS NOT TRUE, UNLESS…

FEDERATED LEARNING ARCHTECTURES REQUIRE ADDITIONAL PRIVACY AND INFORMATION SECURITY PROTECTIONS*

1

Privacy Protection at the Client Side

- Anonymization (VEIL.AI)
- Added fake or synthetic data (VEIL.AI)



2

Privacy Protection at the Server Side

- Aggregation / Diff. Privacy (VEIL.AI)
- Secure Multiparty Computation

3

Security Protection for FL Framework

3

Security Protection for the Federated Learning Framework

- Homomorphic Encryption
- Back-door defender

Research with PRINCETON UNIVERSITY

** Ma, Chuan, et al. "On safeguarding privacy and security in the framework of federated learning", IEEE Network (2020)*

- Federation <u>alone</u> is not a solution to privacy problems

VEIL.AI

AFTER ANONYMIZATION – “IS THE DATA GONE NOW?”

ADVANCED ANONYMIZED ROW LEVEL DATA LOOKS THE SAME AND PERFORMS ALIKE THE ORIGINAL (RAW) DATA

Univariate analysis (Target: TenYearCHD)

Methods	Mean accuracy		Mean prediction loss	
	Orig	***Anon***	***Orig***	***Anon***
Logistic Regression*	0.846	0.842	0.154	0.158
Bayesian Ridge Regression	0.838	0.840	0.162	0.160
Gaussian NB	0.820	0.813	0.180	0.187

Multivariate analysis (Target: currentSmoker+TenYearCHD)

Methods	Mean accuracy		Mean prediction loss	
	Orig	***Anon***	***Orig***	***Anon***
Random Forest Classifier	0.836	0.838	0.071	0.073
Decision Tree Classifier**	0.732	0.745	0.150	0.147
K-Neighbors Classifier**	0.725	0.715	0.133	0.139

AutoML:
Train on Anonymized → Test on Real data results using holdout data. (Target: TenYearCHD)

Mean accuracy		ROC_AUC (area under the curve)		Log prediction loss	
Orig	***Anon***	***Orig***	***Anon***	***Orig***	***Anon***
0.831	0.836	0.720	0.718	0.403	0.401

- **Using automatic optimization of the anonymization process**
- **Summary: VEIL.AI row level anonymized data <u>looks the same and performs alike the original (raw) data</u>**

** published model for Framingham dataset on kaggle.com*
*** modified model for multivariate prediction*

VEIL.AI

DIFFERENTIAL PRIVACY STANDARD vs. VEIL.AI TECHNOLOGY

Multivariate analysis (Target: currentSmoker+TenYearCHD)

Methods	Mean accuracy		
	Orig	***Anon baseline***	***Anon optimized***
Random Forest Classifier	0.844	0.686	0.842
Decision Tree Classifier* **	0.734	0.562	0.761
K-Neighbors Classifier* **	0.725	0.444	0.684

Significant difference $p < 0.05$

Significant difference $p < 0.05$

The difference is not significant $p > 0.05$

** published model for Framingham dataset on kaggle.com*

*** a correlation coefficient value between -1 and +1 where +1 represents a perfect prediction, 0 an average random prediction and -1 an inverse prediction*

**** modified model for multivariate prediction*

ADVANCED ANONYMIZATION OFFERS SUPER QUALITY!

VEIL.AI

SUMMARY

1. Next generation AI based anonymization technology beats “old state-of-the-art” anonymization technologies

Feature	Unique VEIL.AI technology	"Old" state-of-the-art anonymization
Data security and privacy	✓	✓
Data with high quality and utility	✓	X
High performance and scalability	✓	X
Secure multiparty anonymization	✓	X
Multi-modal data protection	✓	X
Continuous and real-time anonymization	✓	X
AI-assisted privacy risk assessment	✓	X
Adaptive anonymization and data synthesis	✓	X

2. Advanced anonymized row data looks the same and performs alike the original (raw) data
    - Advanced anonymized datasets can be used to draw the same conclusions as from the original data

Methods	Mean accuracy		
	Orig	*Anon baseline*	*Anon optimized*
Random Forest Classifier	0.844	0.686	0.842
Decision Tree Classifier* **	0.734	0.562	0.761
K-Neighbors Classifier* **	0.725	0.444	0.684

Significant difference $p < 0.05$

Significant difference $p < 0.05$

The difference is not significant $p > 0.05$

3. Federation is great, but it needs anonymization to solve the privacy protection problems

Research with PRINCETON UNIVERSITY

**** Ma, Chuan, et al. “On safeguarding privacy and security in the framework of federated learning”, IEEE Network (2020)***

VEIL.AI

FOR MORE INFO, PLEASE VISIT:
www.veil.ai

THANK YOU!

CONTACT INFO:

Tuomo Pentikäinen
CEO
+358 40 183 2881
tuomo.pentikainen@veil.ai

VEIL.AI

Expert.ai NLU technology

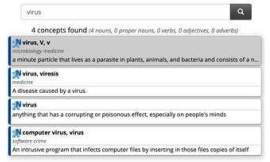
- **Gianluca Sensidoni,** EMEA Account Manager, Expert.ai

- **gsensidoni@expert.ai**
- **+39 3356595360**

What Expert.ai does

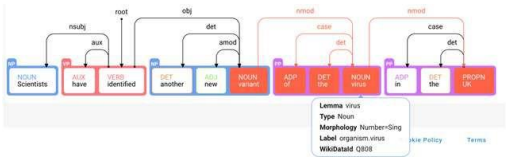
The ability to understand language and transform it into insights

Knowledge Graph



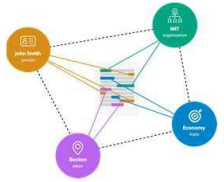
Expert.ai's knowledge graph is a representation of the real world where concepts are defined and connected to one other by semantic relationships.

Natural Language Understanding



Expert.ai distinguishes the correct **meaning of words** and expressions in context and automatically associates the attributes of more general terms that are conceptually linked to these words.

Reasoning Process



As the accurate comprehension of textual content always requires domain-specific knowledge. **Hybrid** Expert.ai technology can **mimic human** reasoning while analysing either on-line or off-line contes

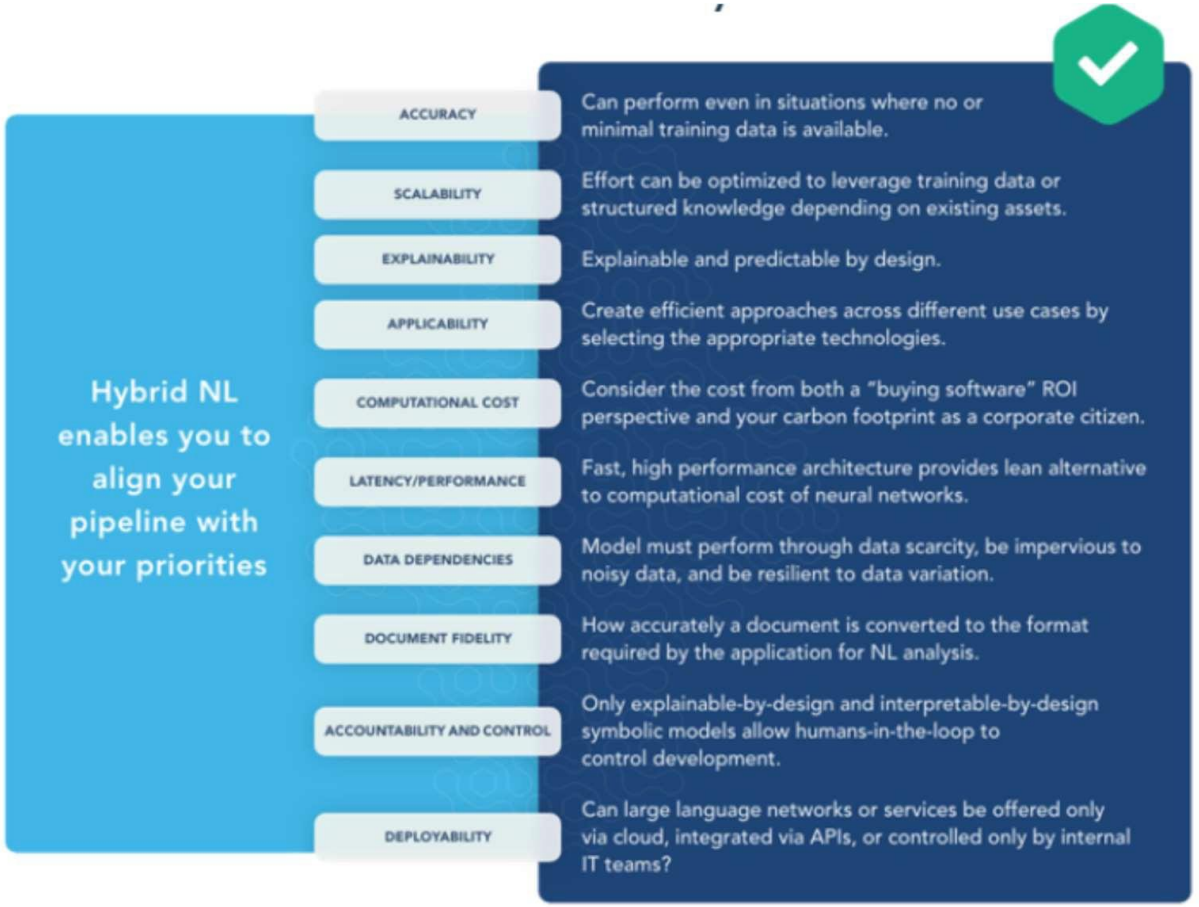
YOUR EXPERTISE AT SCALE

Embedded Knowledge

Disambiguation

Actionable Insight

The benefits of Hybrid NL



Expert.ai Technology …. at the end

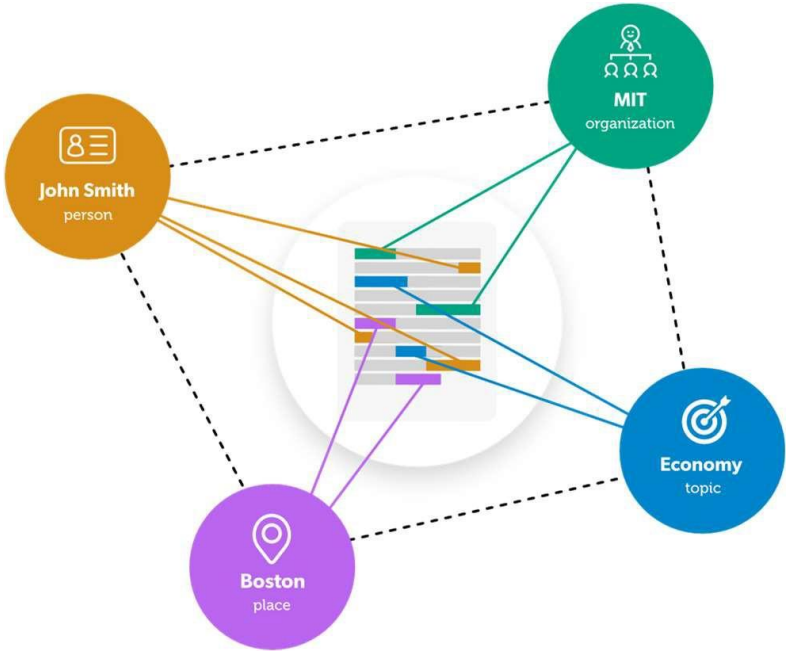
Expert.ai uses **patented AI algorithms** to mimic the human ability to read and understand any textual content

Expert.ai is based on a **pre-trained Knowedge Graph** with millions of concepts, entities, synonyms in **15 different languages**

Unlike traditional ML, it **does not require thousands or millions of examples** to be "trained from scratch"

Uses **advanced semantic rules** to classify and extract information with one of the higher accuracy levels on the market*

It's **not a black-box**: its Explainable AI concepts can always be described, documented and audited



How to use it

The target is to extract semantic data from unstructured contents in order to gain ***Information Superiority*** useful for ***Augmented DSS***, ***Simulation & Prediction*** environment.

In the Healthcare domain f.e:

- Clinical Research
- Medical Report & Patient records analysis and cross correlation
- Disinformation vs Misinformation highlighting (Cognitive Warfare)
- Pandemic Surveillance
- Social Impact Discovering
- Digital Disease Detection
- Pharmacovigilance

INSERM in the Clinical Research

How do you detect and act on clinical insights when they are hidden across 100 million medical documents?

The Challenge

Overwhelming volumes of medical information (scientific literature, clinical studies, patents, side effect reports, and news) make it increasingly difficult for life sciences professionals to stay ahead of and develop relevant insights into their scientific and competitive landscape

How AI helps

Analyze 300+ information sources (Scientific and Healthcare)

Extract and Map connections between molecules, mechanisms of action, Pharmas & Biotechs, side effects, clinical trials, research, and networks of experts

The Impact

Days/month

Time Savings

Professionals can focus on analysis & decisions instead of data gathering

About Expert.ai

€ 32m
Revenues

~270
employees

Public Company (EXSY)

- Largest global company specialized in Nybrid Natural Language Understanding technology
- Direct branches in US, CAN, UK, FR, DE, ES & IT

Dedicated to Customer Success

- Award winning Patented technology
- 30 years of R&D investments to create expert.ai technology
- Local technical & client teams committed to our customers' business & technology needs

Award Winning

2020 Best Overall NLP Company

Magic Quadrant for Insight Engines

AI-based Text Analytics Platforms

Industry Recognition

FORRESTER®

AI-Based Text Analytics

People & Document-Focused Platforms

"…expert.ai's knowledge graph beats purely ML-based text analytics apps …achieving faster time-to-value, more predictable results, and lower overall cost of ownership."

2020 Forrester Wave

FINTECH REGTECH Global Awards 2020

Machine Learning and Artificial Intelligence Partner of the Year

AI BREAKTHROUGH AWARD 2020 BEST OVERALL NLP COMPANY

Best Overall Natural Language Processing Company

100 COMPANIES That Matter in Knowledge Management 2020

KMWorld 100 Companies that matter in KM

Live Demos

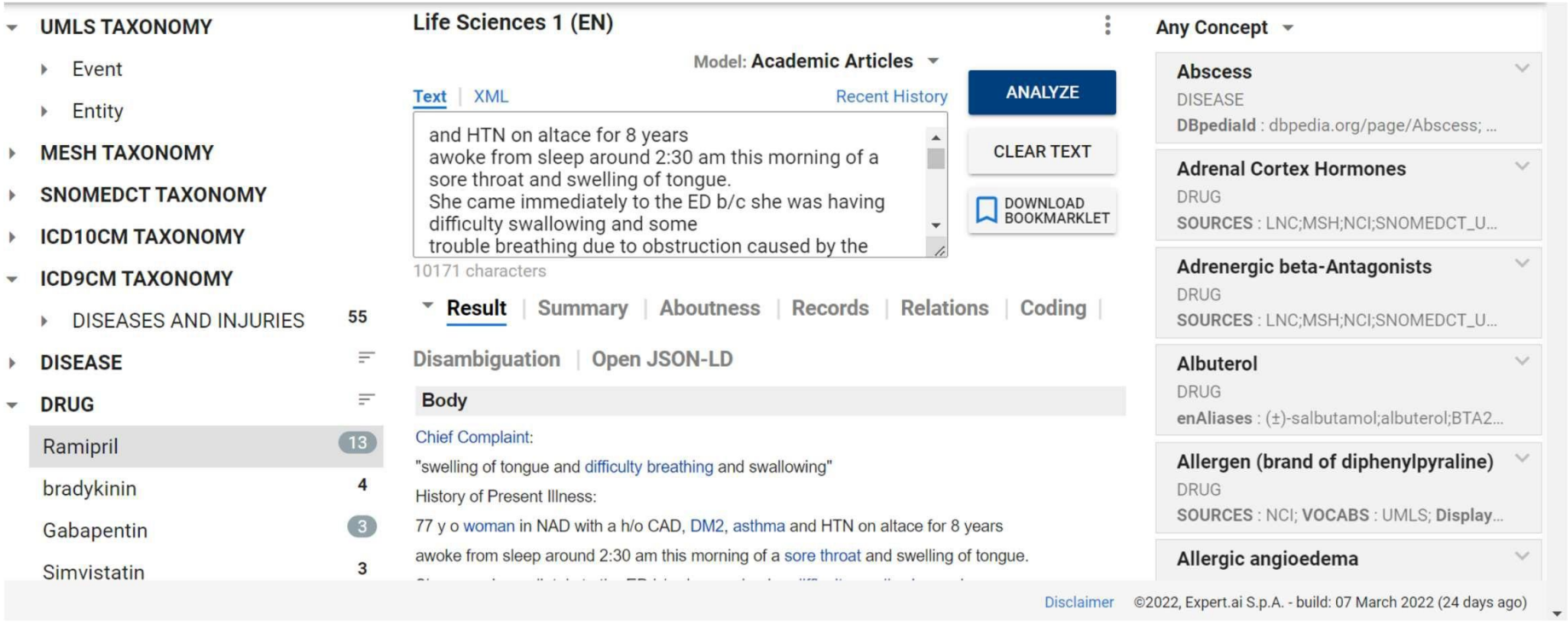
Enjoy our **Free Live Demos** at:

- https://try.expert.ai - standard and available in 5 languages
- https://www.intelligenceapi.com/demo/ - available in English language and dedicated to Stylometric and Emotional analysis (Behavioural algorithms for Human Factor attributes extractions)

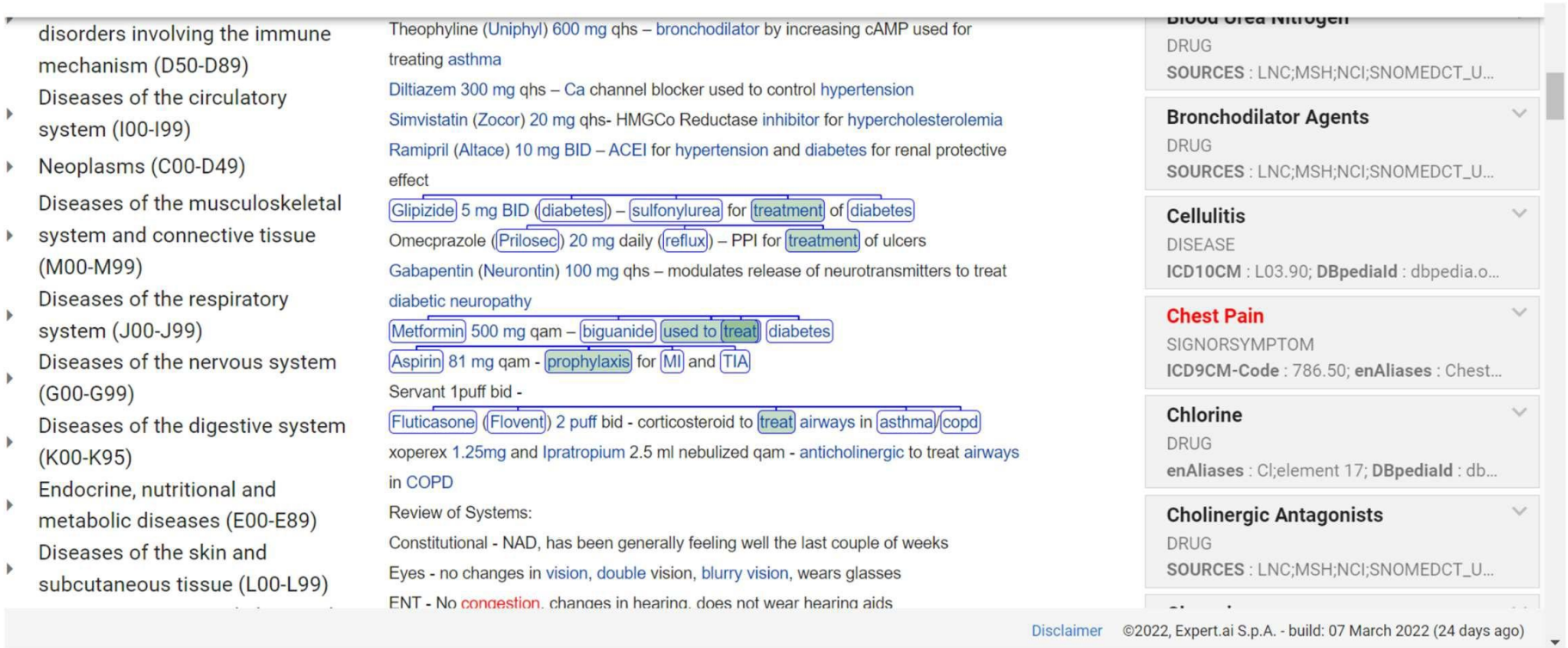
Healthcare Live Demo with domain oriented international vocabularies such as ICD-9, ICD-10, UMLS, SNOMED, MESH, IUPHAR and ChemID*plus*.

(please ask me (gsensidoni@expert.ai) for a dedicated session)

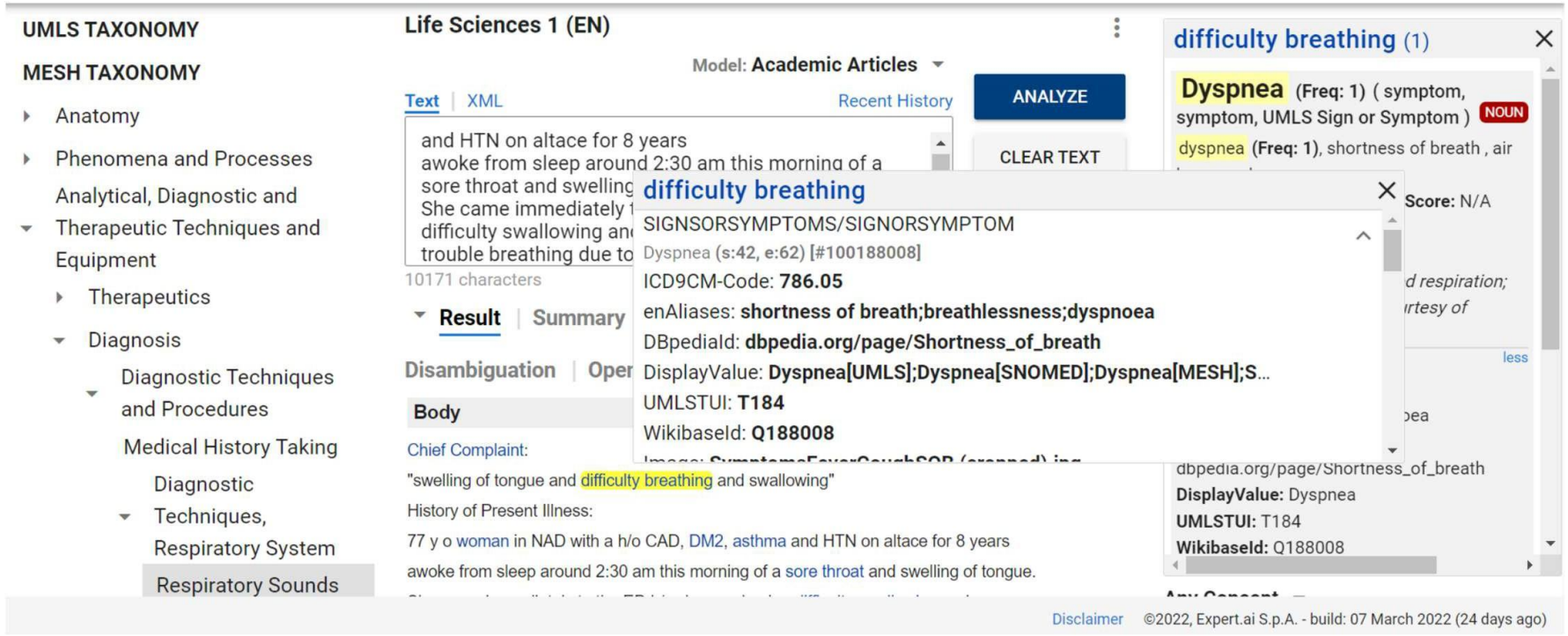
Healthcare Live Demo (1)



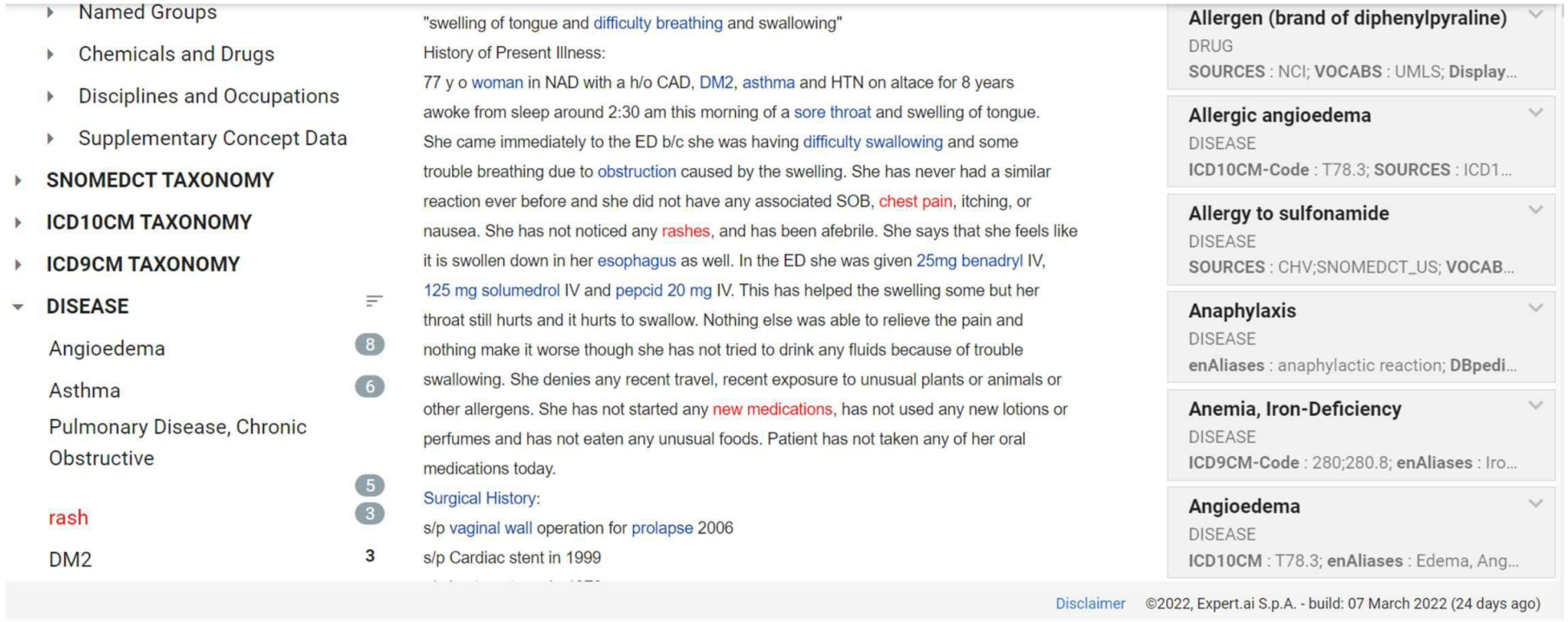
Healthcare Live Demo (2)



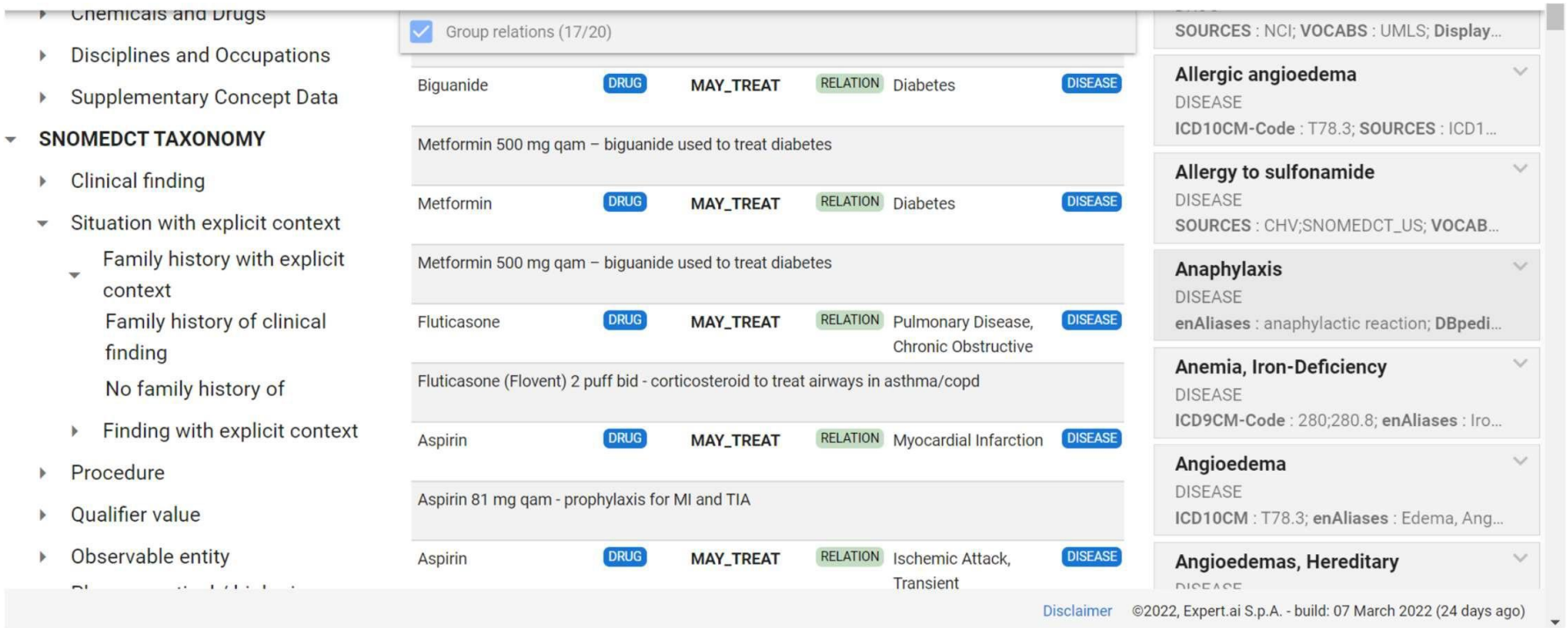
Healthcare Live Demo (3)



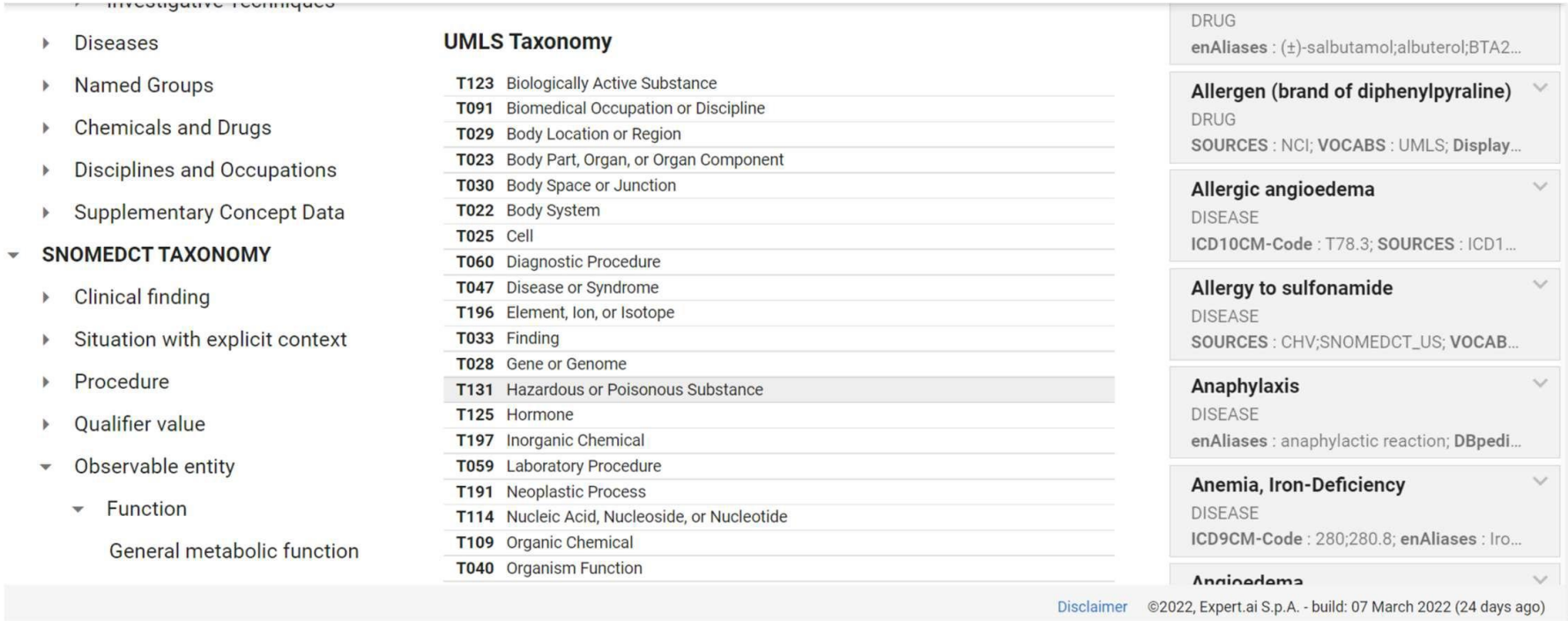
Healthcare Live Demo (4)



Healthcare Live Demo (5)



Healthcare Live Demo (6)



AI-Evidence – AIR_PTE

Prof. Dr. Thomas P. Zahn
DCC Risk Analytics & DSI Data Science Institute

thomas.zahn@risikoanalytik.de

+49-30-3199829-01

AIR_PTE
AI based Risk Prediction & Treatment Effect Estimation

International Collaboration Project funded by BMWI & IRAP

- Current Partners from Germany, Canada, Denmark and Romania – open to join now

Goal: Rapid Evidence Generator (REG)

Risk Adjusted Propensity Score (RAPS) using gradient boost & deep Learning

- on historic (spatio-temporal) treatment pattern (ICD , ATC , OPS …)
- to identify outcome risk adjusted control for multiple interventions in real world claims data

Applicable to generate RCT comparable Evidence – fast, reliable and effective i.e. for

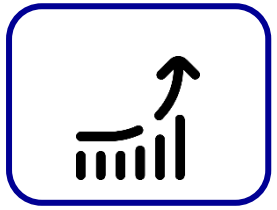
- Innovative treatment options (DOAC for VTE treatment study) & programs (Smart CasaPlus)
- Digital Health Applications (DiGA)

Gefördert durch: Bundesministerium für Wirtschaft und Energie aufgrund eines Beschlusses des Deutschen Bundestages

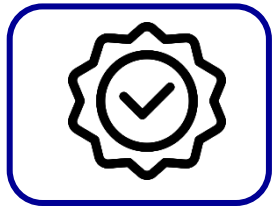
AIR_PTE
AI-Evidence

Problem

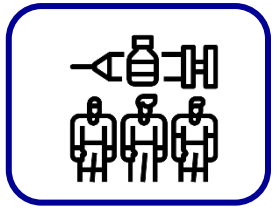
gaia-x



Health **innovation speed** is exponentially increasing i.e. in pharma, devices and digital health



Innovations in Health need **approval** by regulators **based on scientifically validated studies**

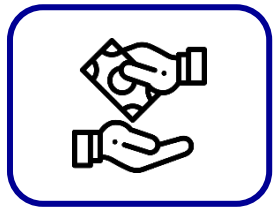


The Gold Standard - randomized control trials (**RCT) – are very time consuming and expensive**

Real World Application of Health Innovations often **differ from RCT study populations**

HENCE

Innovators need Real World Evidence to **continuously monitor effectiveness and efficiency** of their innovations



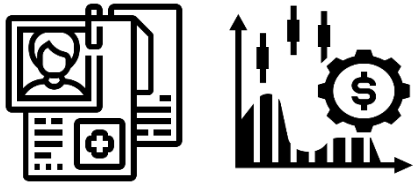
Public Healthcare Systems need Real World Evidence to **focus Innovations** to populations with highest benefits

Medical Practitioners need Real World Evidence to **support personalized treatment decisions**

AIR_PTE
AI-Evidence

Solution

gaia-x



Public Healthcare Systems regularly produce **long term health claims data**

mirror the real-world effect of innovations **on the entire population**

can be **used for the purpose of generating Real World Evidence**

Real World Evidence from claims data is **fast, inexpensive and continuously** obtainable

BUT

contain highly sensitive personal information and **need the highest level of data protection**

are obtained for the purpose of reimbursement and **not always available for research**

are **highly structured and standardized** but **may lack specific information typical to RCT**

HENCE

smart AI algorithms are needed

to **ensure full data protection** when using health claims for research

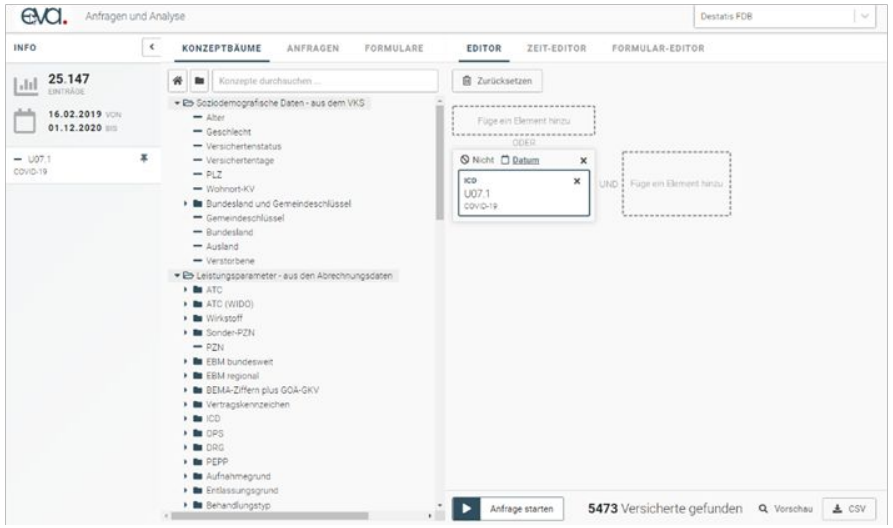
for the retrospective **identification of study sub-populations and outcomes**

for the **identification of specific intervention and valid control groups**

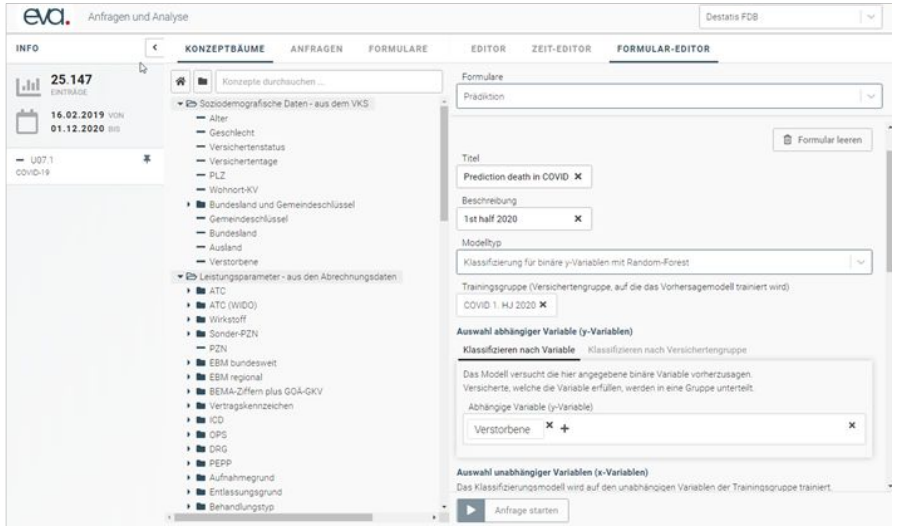
for generating **Real World Evidence for innovative Treatment** options

we aim to adapt the REG Methods to the needs of COVID-19 Pandemic Risk predictions

1.) Approx 5500 patients with COVID from 1st wave detected



2.) ML algorithm was applied on the COVID cohort to predict death



World Health Organization

Health Topics | Countries | Newsroom | Emergencies

Home / News / WHO, Germany launch new global hub for pandemic and epidemic intelligence

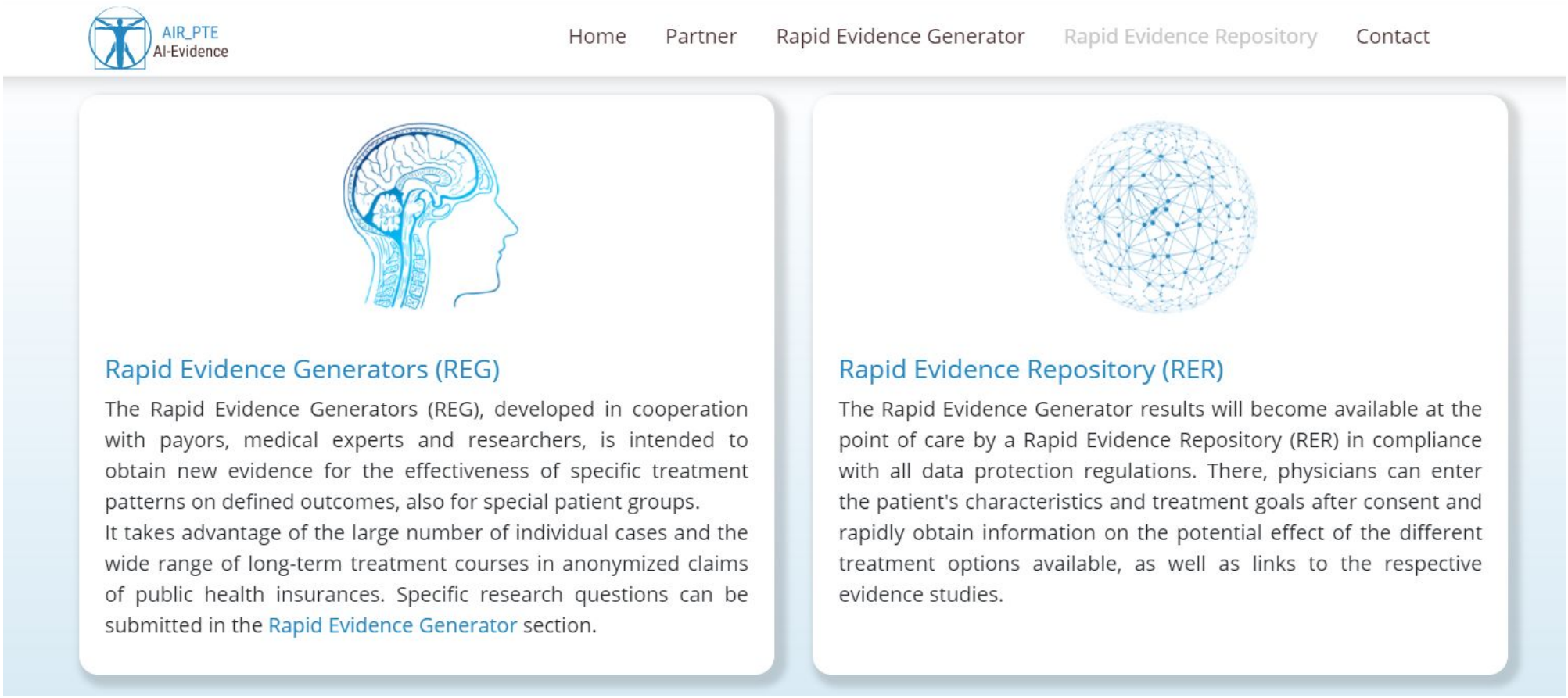
WHO, Germany launch new global hub for pandemic and epidemic intelligence

5 May 2021 | Joint News Release | Geneva/Berlin | Reading time: 2 min (567 words)

AIR_PTE
AI-Evidence

Together, we will provide REG Results to Payors, Providers and Researchers via an international Rapid Evidence Repository

www.ai-evidence.de



Contact: Prof. Dr.-Ing Thomas P. Zahn, Vice President Research bbw University, CEO DCC Risikoanalytik Tel. +49 171 7909941 e-mail: thomas.zahn@risikoanalytik.de

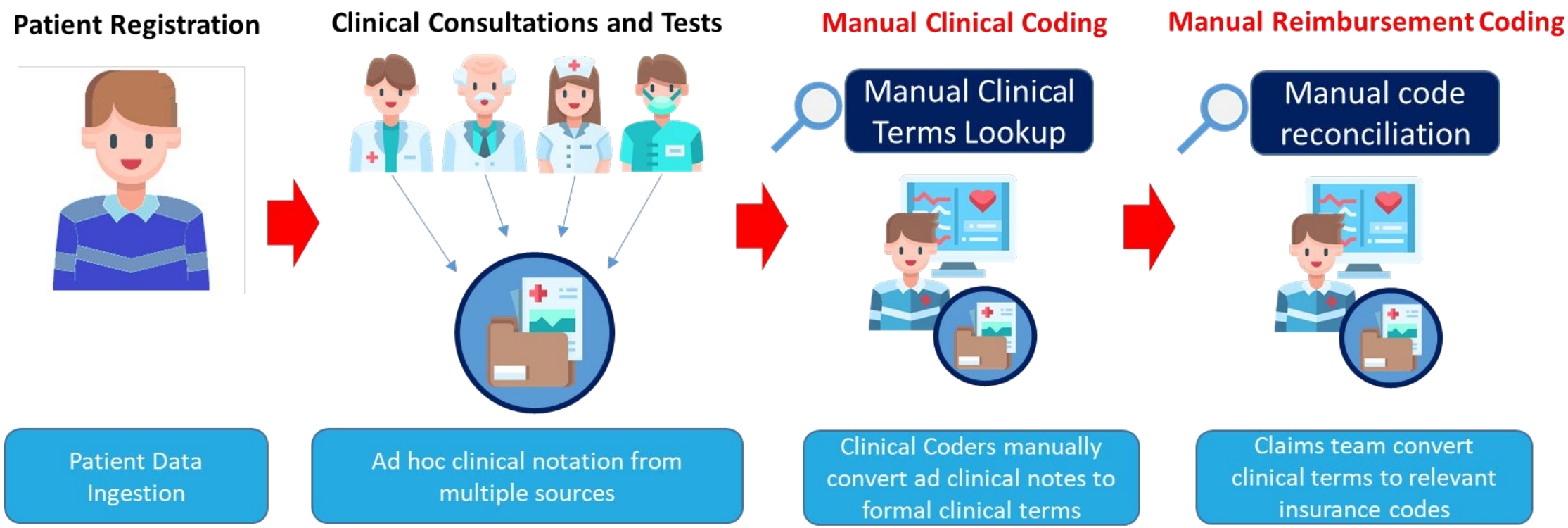
Trusted AI for Data Interoperability

- **Dermot Doyle, CEO**, Dynaccurate

About Us…

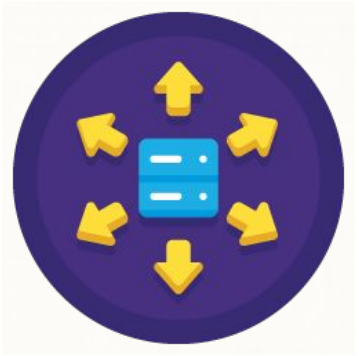
- Dynaccurate is a spin out from Luxembourg Institute of Science and Technology **(LIST)**
- We provide **AI solutions** for health data interoperability
- We use both **Experts Systems**, **Machine Learning** as well as advanced search technologies
- Our mission is to provide **affordable, trustworthy AI** to the health sector to help make health data FAIR
- **The health sector actually requires AI for data management** much more than many other sectors
- Health data sector **underserved for AI solutions**

The reality of health data processing right now…



X Slow - Bottlenecked – Siloed – No Automation **X**

Why so tough??? – Financial Data vs. Health Data



In Finance

- Bank accounts and Bank identifiers rarely change **(Stable)**
- Data is largely numeric and can be worked on with formula, spreadsheets, basic mathematics **(Easier to process with applications)**
- Common on identifiers - IBANs, ISINs, CUSIPs, SEDOLs etc. - and SWIFT infrastructure **(Global regime)**
- Banks motivated to make it happen - More transaction volume, more fees, more profit – (**Common Incentives)**

In Health

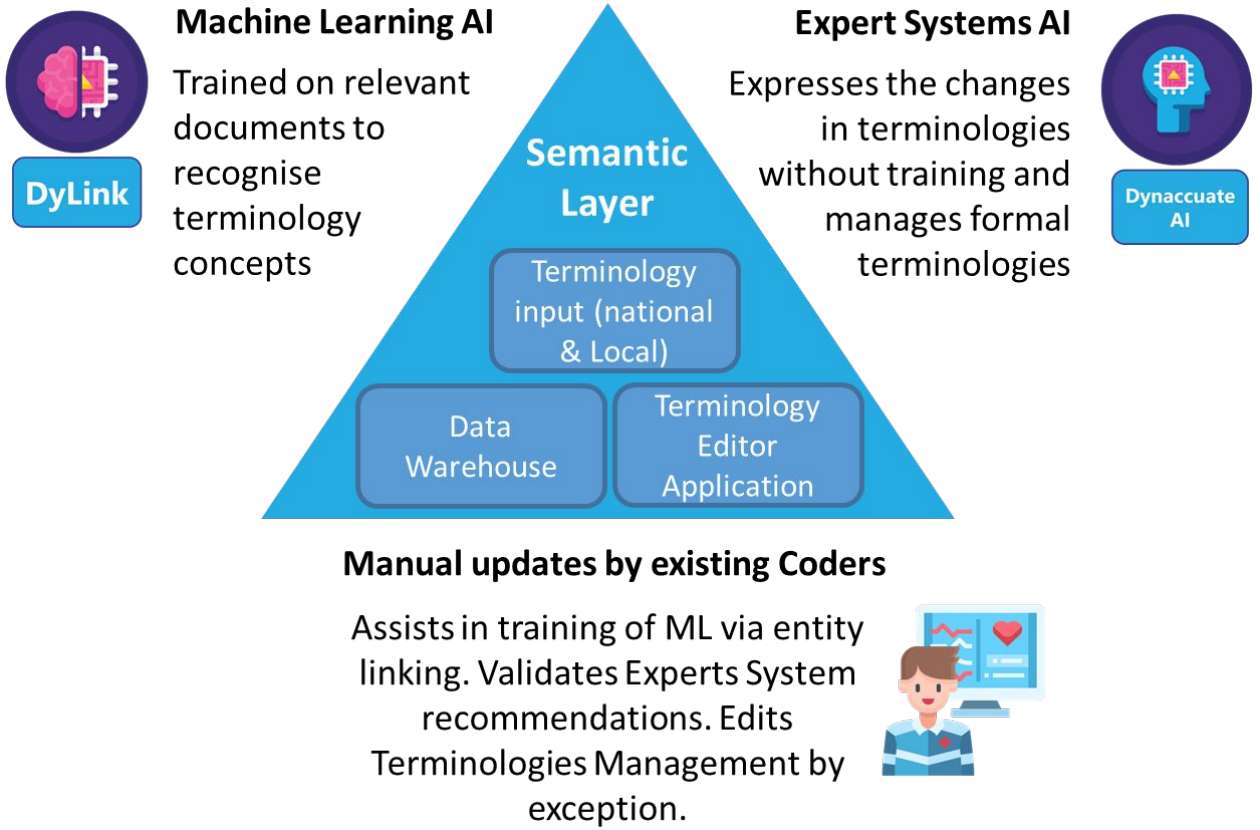
- Health data identifiers like SNOMED, ICD and local terms etc. change a lot **(not stable)**
- Data is largely expert domain terminologies / ontologies and human medical knowledge **(very few applications – requires Semantic Web AI)**
- Common identifiers (i.e. SNOMED, ICD, Global Unique Identifiers) but not backwards compatible **(heavily fragmented)**
- Countries did, and do, 'their own thing' **(few incentives to change, fewer international incentives)**

Semantic Interoperability Triad

Because of the complexity of health data, there's no easy solutions…

AI is fundamental…

Expert Systems and Machine Learning AIs are the main solutions.



Machine Learning vs. Expert Systems

Machine Learning

- Requires data sets & human training
- Quality is determined by both data & training
- When trained can ‘read’ data
- Many off-the-shelf tools
- Expensive to implement
- Not transferrable between languages…

Risks/Quality factors:

- Should be trained on data sets of the type which it will later process.
- Training an AI can be quite mundane and repetitive.
- Will need ongoing training to maintain accuracy

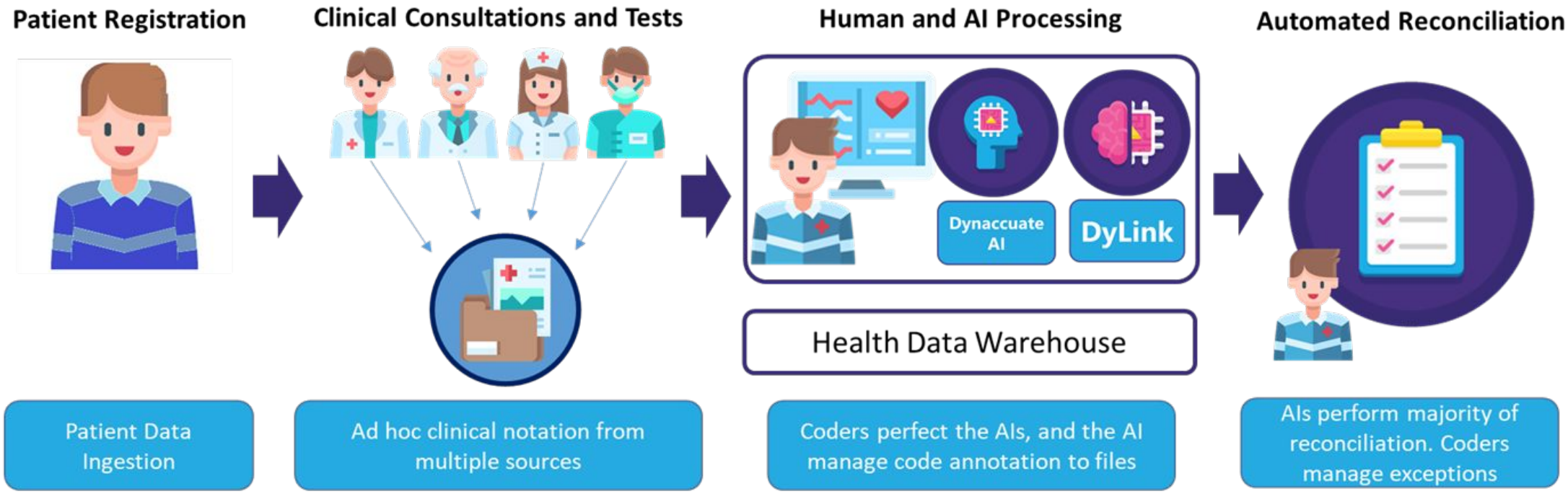
Expert Systems

- No training – the AI simply expresses the changes
- Quality based on sophistication of the rules
- Manages the meta-data
- Much cheaper to implement
- Transferable
- Expensive to build (> €1m!!!)

Risks/Quality factors:

- ‘Low hanging fruit’ has all been taken
- Initial algorithms will need to be persistently tested. Better if peer reviewed
- Depending on how generic the algorithms are, they may transfer or not

Short term impact



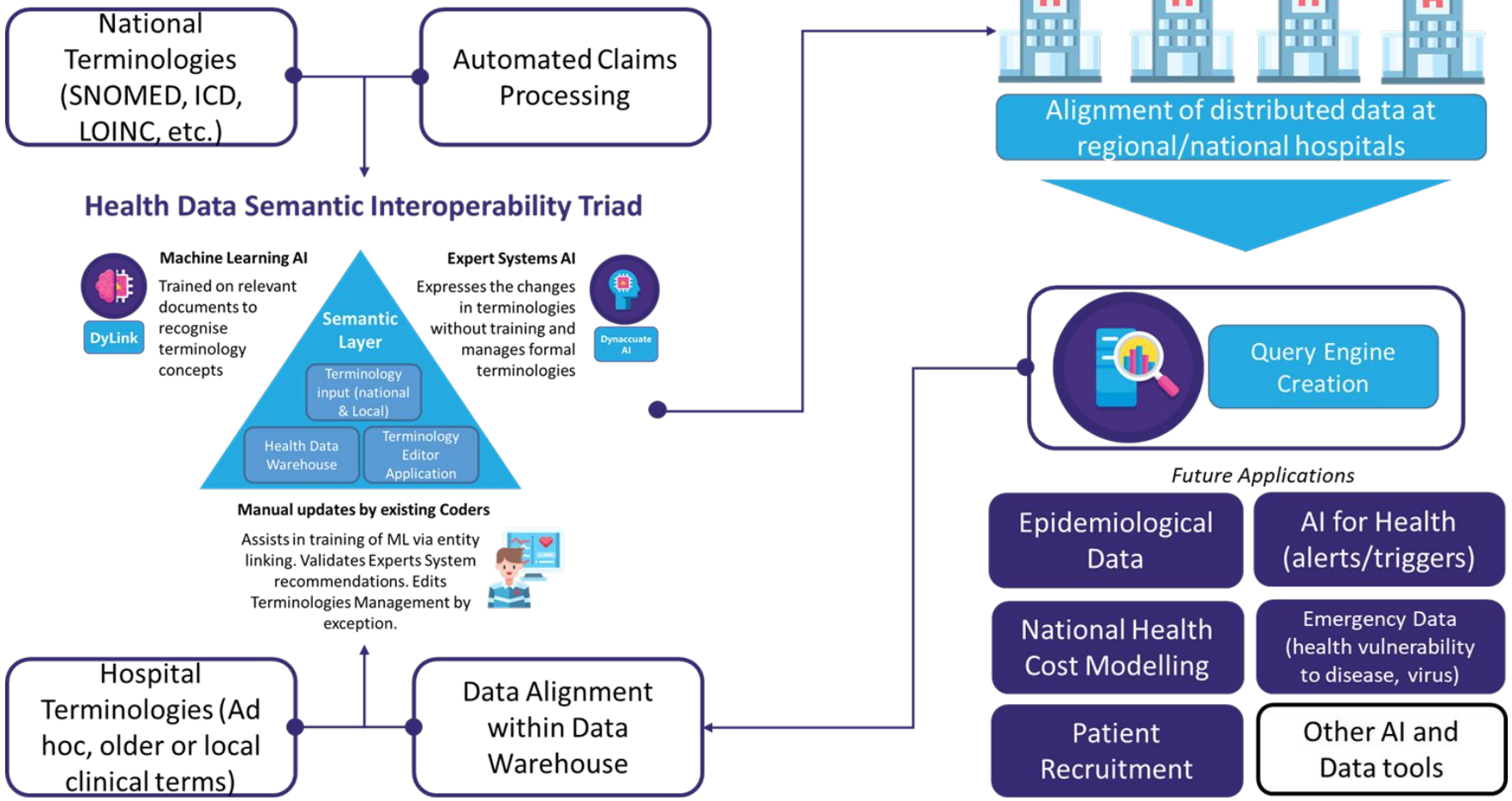
Most clinical coding becomes automated…

Long Term Impact…

gaia-x

As AI becomes prevalent the following will happen:

- **Manual coding disappears**
- Skilled staff work on **AI validation** & perfection
- **Capture of all health indicators** (i.e. symptoms data), not just reimbursable codes
- '**Super AI**' enabled by better health data



Reliable, granular health data enables a completely new economy

Q&A with the audience

gaia-x

Compile session summary