Harnessing Unstructured Data & Hospital Interoperability

The Potential of Next Generation Platform For Transforming Patient Recruitment

---

OHDSI Sweden Symposium 2025, Stockholm, November 5th, 2025

Karolinska Institutet

OMOP4SWEDEN!
EVIDENCE GENERATION FOR BETTER HEALTHCARE

"From data silos to learning systems in clinical research."

Mats Sundgren, PhD

Contents

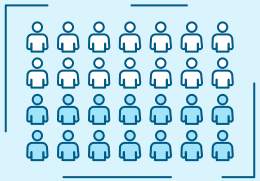
(Still) Problems with Clinical Trials

Incomplete and delayed clinical trials are a sore spot of drug development

The percentage of studies that fails to meet enrolment target:

37 % **sites under-enroll**

11 % **Zero patients**



80%

of today's clinical trials fail to achieve the target recruitment

Almost

40%

Recruitment can account for 40 % of trial budget [2]

Onco Phase III trials now exceed $36 million, up 30 % since 2018.

$36 m

Each day of delay in drug development costs sponsors on average up to $0.8m in lost revenue [3]

$0.8m

1. Clinical Trials Statistics 2025 By Phases, Definition and Interventions.
2. Tufts 2024

Data First, AI Later

- **Integrity and Integration of Data:**
  Data recognized as the critical foundation of successful digital health initiatives.
  Importance of the "3i's of data": Integrity, Integration, and Intelligence (AI) stressed.

- **Policy and Education:**
  Limited data literacy among policymakers poses a barrier to the effective use of AI.
  Calls for better education to facilitate responsible AI deployment.

- **Privacy and Security Concerns:**
  Highlighted risks of implementing AI prematurely without robust data frameworks.
  Emphasized the necessity of addressing privacy and security issues early on.

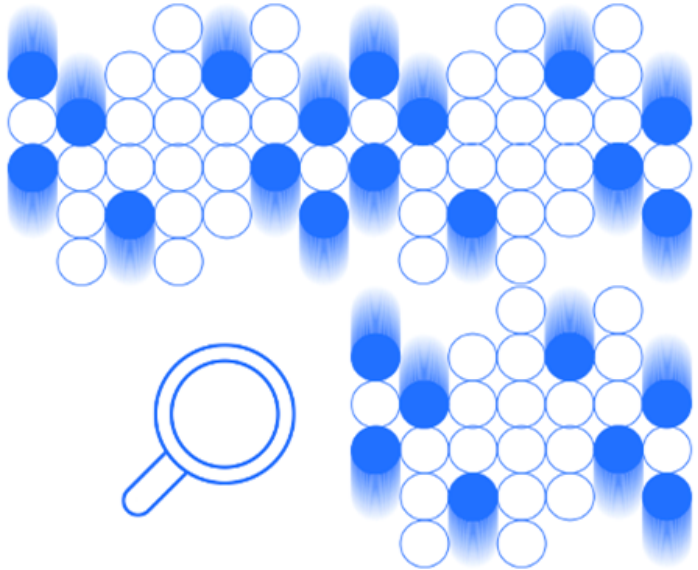
- **Harmonizing Regulations:**
  Discussed the varying data privacy laws between regions.
  Expressed optimism for the future convergence of regulatory standards, especially in managing health data.

Reference: Data First, AI Later: Steering Towards a Data-Centric Future in Healthcare, Health Data Forum, Geneva 2024

Digitized patient centric recruitment
The case of IOMED

Context – The Future of Healthcare

2.3 Exabytes
of Real World Data in healthcare are **not fully leveraged.**



Unstructured information.

65-85% of patients' clinical information stored in hospitals is written in unstructured text format.

Siloed information.

Data is still disperse and mostly accessible by only hospital by hospital.

Zhao X et al. Integrating real-world data to accelerate and guide drug development: A clinical pharmacology perspective. Clin Transl Sci. 2022 Oct;15 (10):2293-2302.

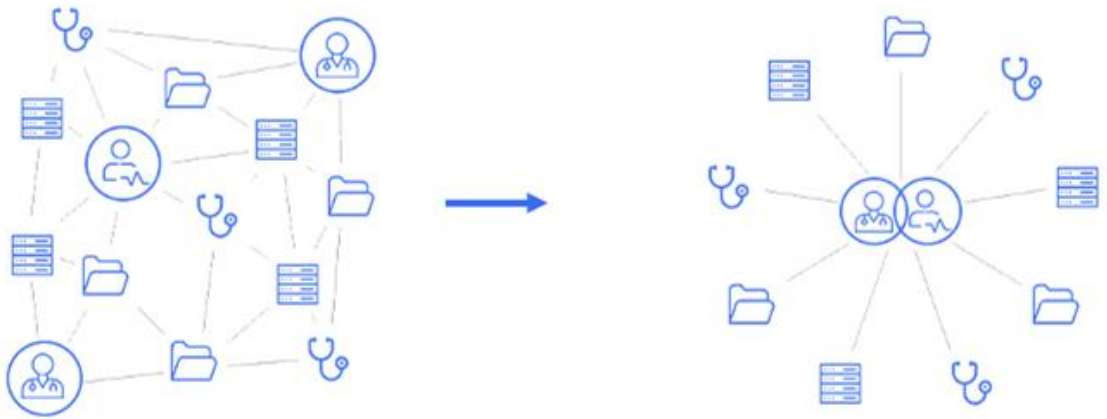
Courtesy: IOMED

Patient-centric approach

Digitization of health data has been **process-centric**, making data messy, sparse and inaccessible.

We make data **patient-centric**

We bring order to **Real World Data thanks to AI.**



Legacy Healthcare Data Architecture

Patient-centric Data Architecture

Courtesy: IOMED

How data is standardized and made ready for research use

Data ready to be consumed thanks to **Artificial Intelligence**

1 Automation of Extraction:

AI enables the automatic collection of large volumes of data, freeing up time and human resources.

2 Automation of Normalization:

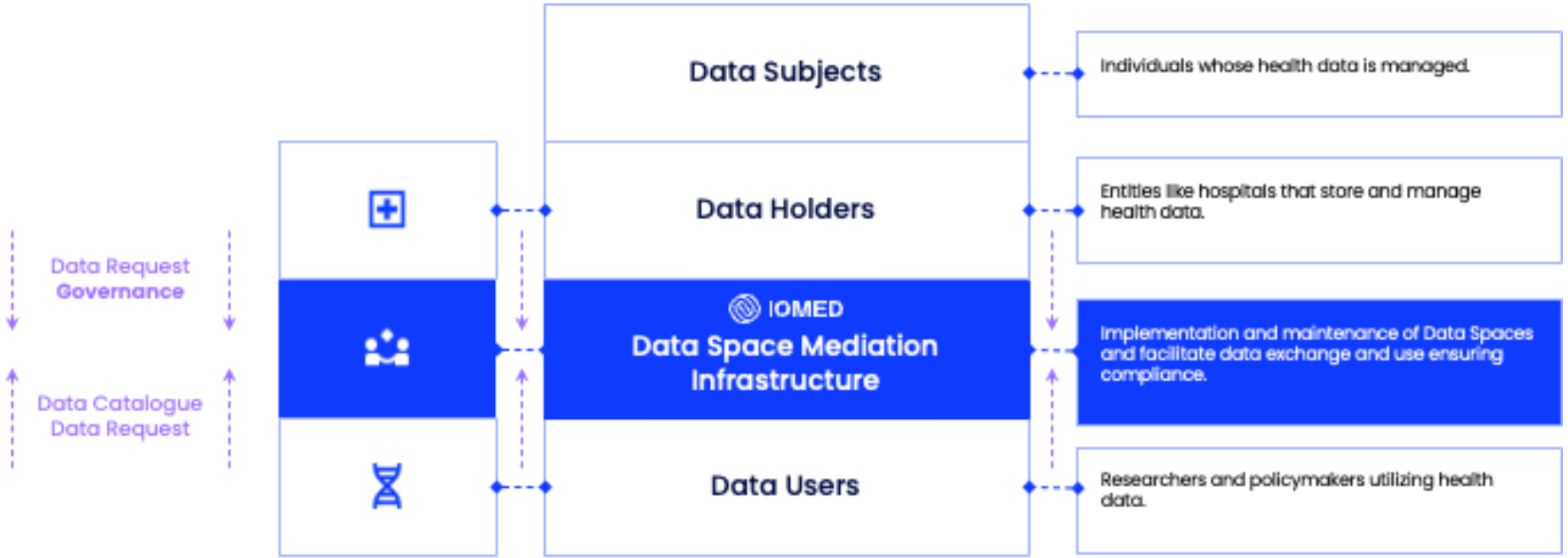
Large Language Models based *Mapping* and *NLP* facilitates the normalization of data to standards, ensuring consistency and compatibility.

3 Data quality at Scale:

AI algorithms can identify and correct errors or inconsistencies in data, improving its quality.

Courtesy: IOMED

Data ready to be consumed & how standardize data



Courtesy: IOMED

Data Space Platform evidence

PATIENT IDENTIFICATION

Identification of patients candidates to Cellular therapies for Lymphoma.

The project is helping clinicians identify more patients that are candidates to Cellular therapies. The narrow time window for therapy implementation between the moment the conditions are met and they are candidates for the treatment.

ONGOING

Fast Track Review of Patients by Clinicians
An 85% of Accuracy

EXPECTED PATIENTS	3.3x	PATIENTS
3		10

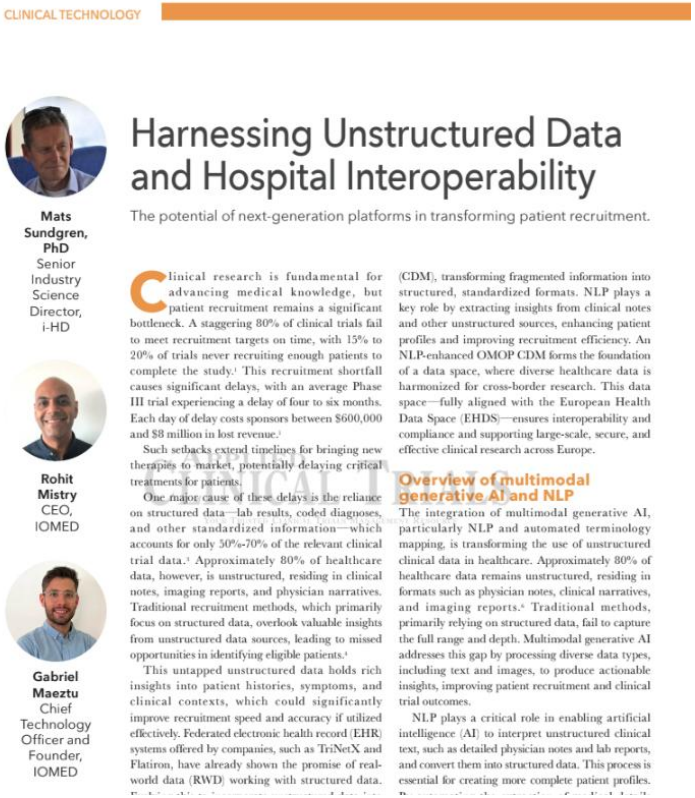
Patient selection criteria

a/ Diagnosis of relapsed or refractory diffuse large B-cell lymphoma (DLBCL)

b/ Diagnosis of relapsed or refractory primary mediastinal large B-cell lymphoma after at least two previous systemic therapies (PMLBCL)

c/ Diagnosis of refractory acute B-cell lymphoblastic leukemia (LLA B)

Data Space Platform



CLINICAL TECHNOLOGY

Harnessing Unstructured Data and Hospital Interoperability

The potential of next-generation platforms in transforming patient recruitment.

Mats Sundgren, PhD
Senior Industry Science Director, i-HD

Rohit Mistry
CEO, IOMED

Gabriel Maeztu
Chief Technology Officer and Founder, IOMED

Clinical research is fundamental for advancing medical knowledge, but patient recruitment remains a significant bottleneck. A staggering 80% of clinical trials fail to meet recruitment targets on time, with 15% to 20% of trials never recruiting enough patients to complete the study.[1] This recruitment shortfall causes significant delays, with an average Phase III trial experiencing a delay of four to six months. Each day of delay costs sponsors between $600,000 and $8 million in lost revenue.[2]

Such setbacks extend timelines for bringing new therapies to market, potentially delaying critical treatments for patients.

One major cause of these delays is the reliance on structured data—lab results, coded diagnoses, and other standardized information—which accounts for only 50%-70% of the relevant clinical trial data.[3] Approximately 80% of healthcare data, however, is unstructured, residing in clinical notes, imaging reports, and physician narratives. Traditional recruitment methods, which primarily focus on structured data, overlook valuable insights from unstructured data sources, leading to missed opportunities in identifying eligible patients.[4]

This untapped unstructured data holds rich insights into patient histories, symptoms, and clinical contexts, which could significantly improve recruitment speed and accuracy if utilized effectively. Federated electronic health record (EHR) systems offered by companies, such as TriNetX and Flatiron, have already shown the promise of real-world data (RWD) working with structured data. (CDM), transforming fragmented information into structured, standardized formats. NLP plays a key role by extracting insights from clinical notes and other unstructured sources, enhancing patient profiles and improving recruitment efficiency. An NLP-enhanced OMOP CDM forms the foundation of a data space, where diverse healthcare data is harmonized for cross-border research. This data space—fully aligned with the European Health Data Space (EHDS)—ensures interoperability and compliance and supporting large-scale, secure, and effective clinical research across Europe.

Overview of multimodal generative AI and NLP

The integration of multimodal generative AI, particularly NLP and automated terminology mapping, is transforming the use of unstructured clinical data in healthcare. Approximately 80% of healthcare data remains unstructured, residing in formats such as physician notes, clinical narratives, and imaging reports.[4] Traditional methods, primarily relying on structured data, fail to capture the full range and depth. Multimodal generative AI addresses this gap by processing diverse data types, including text and images, to produce actionable insights, improving patient recruitment and clinical trial outcomes.

NLP plays a critical role in enabling artificial intelligence (AI) to interpret unstructured clinical text, such as detailed physician notes and lab reports, and convert them into structured data. This process is essential for creating more complete patient profiles.

"Unlocking value from unstructured data requires natural language processing, a robust data governance framework, and a clear focus on delivering value to healthcare stakeholders."

The study highlights the vast potential of unstructured health data—such as clinical notes, radiology reports, and lab results—which make up around 80% of all medical information generated.

Traditionally hard to analyze, this data can now be unlocked through advanced natural language processing (NLP) techniques like those used by IOMED.

The company focuses on converting this rich but underused resource into structured, standardized data that fuels clinical research and evidence-based decision-making, all within a strong governance framework that safeguards privacy.

IOMED Milestones

Courtesy: IOMED

Outlook & Conclusions

Transforming Trials with AI, eSource & RWD

1. From silos to synergy: Hospitals, sponsors, and innovators are aligning on open standards and federated data models.

2. Data first, AI later: OMOP-based, high-quality clinical data is the foundation for trustworthy and scalable AI.

3. Unstructured data → recruitment power: NLP and LLMs reveal trial-eligible patients hidden in clinical text, bridging structured and unstructured sources.

4. Federated collaboration: Insights are analyzed locally, shared globally, preserving privacy while building confidence in results.

5. The path forward: Continuous validation, transparency, and OMOP-driven interoperability will make AI-enabled, patient-centric trials the new normal.

"OMOP as the backbone — AI as the amplifier — collaboration as the catalyst."

Thank You

Scaling up the technology opens new avenues for collaboration!

This is exemplified by deploying the output from new innovative solutions which involves pharmaceutical companies, hospitals, along with new vendors and academic partners.

It serves as a prime example of a **Nash equilibrium**

- the best results emerge when each participant in the group acts in their individual best interest, contributing to the collective success of the entire group.

(Governing dynamics 1962 - John Nash, Nobel Laureate in Economics, 1994)

i~HD is a neutral not for profit body, bringing stakeholders together