

# ICA ROMA 2022



September 19<sup>th</sup>-23<sup>rd</sup>, Italy



## A New Era for Health Research by Using AI

# ARCHIVES: BRIDGING THE GAP

9<sup>th</sup> Annual Conference of the  
International Council on Archives

19-23 September 2022 | Rome, Italy

Kjersti Aasgård

Gina Helstad



The Norwegian Health Archives Registry

# A New Era for Health Research by Using AI

Results from a pilot study demonstrating how AI can analyze  
historical patient records

Gina Helstad and Kjersti Aasgård





ARKIVVERKET  
NORSK HELSEARKIV

# Introduction

- › Part of the National Archives of Norway
- › Established in 2019
- › Two purposes:
  - › Making health data available for research
  - › Securing and preserving patient records from deceased patients
- › 36 million digitized pages pr year
- › All hospitals are required by law to transfer all patient records to The Norwegian Health Archives after the patient is dead
- › First national health archive in the world



“

**To become a  
recognized health  
registry for new  
knowledge and  
better public health**

---

Strategic goal





## INPUT:

- 15 million paper-based patient records
- 150.000 shelf meters in total
- Electronic patient records: backlog of 2 million

## PROCESS:

- Digitization and disposal of the paper
- Metadata registration:
  - Personal information
  - Diagnoses
- Digital preservation system
- OAIS compliant

## OUTPUT:

- Digital access
- Statistics and complete records (images and OCR-text)
- Access only by application
- Researchers and next of kin








## The Norwegian Health Archive Registry

Content overview September 2022


 **700 000 patient records**

 **53 hospitals**

 **3 700 000 diagnoses**

 **1871 – 2016: current time span**

 **244 million files**

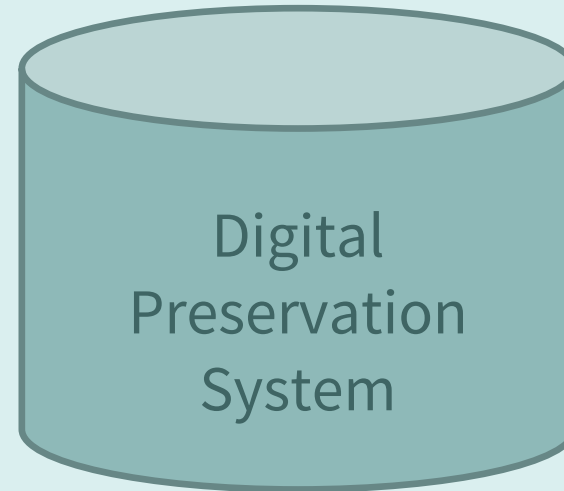
 **100 TB data**

 **85 million pages digitized**

## The Health Archives Registry



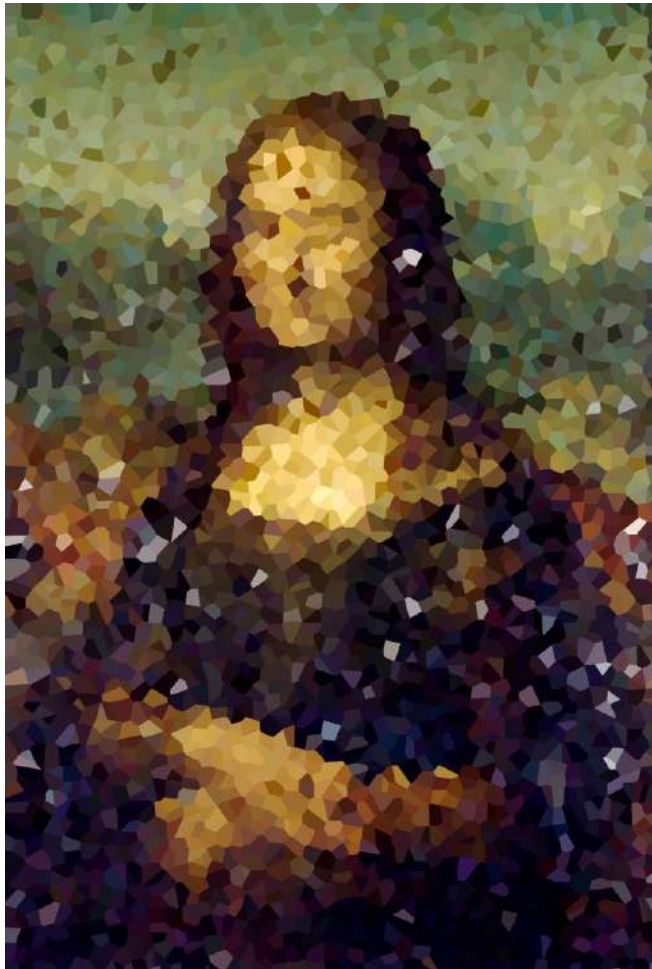
Metadata  
Patient ID, diagnoses  
*structured*



Images – digital format  
OCR text  
*unstructured*



# The Health Archives Registry contains both **structured** and **unstructured data**



- › The **structured** data are systemized, «tidy», easy to use and well suited for statistical analysis
- › The **unstructured** data can be more challenging to handle and access, but that does not mean the data are less valuable



# Klokketest

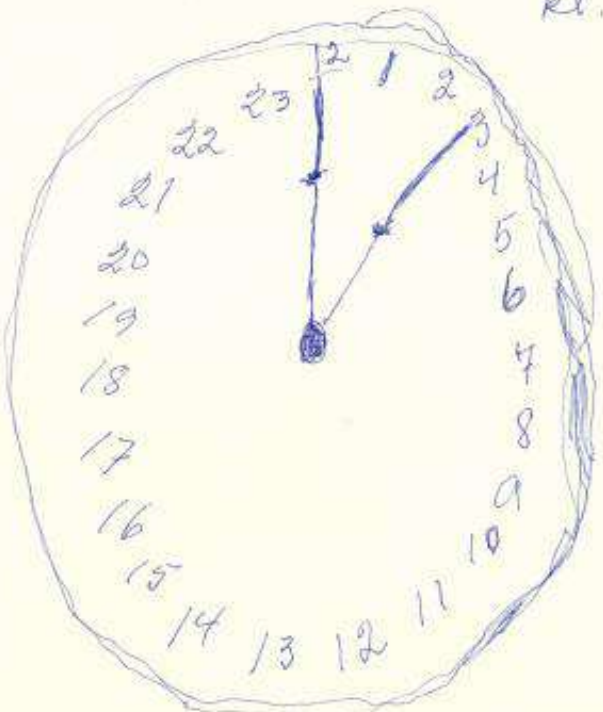


### Instruksjoner:

På feltet under, vennligst tegn en urskive og plasser alle tallene på korrekt sted.

Etter det, tegn inn viserne slik at klokken viser ti minutter over elleve.

kl. 3<sup>00</sup>











# Innovation partnership with Anzyz Technologies

- › Use artificial intelligence to find relevant information in big data
- › Test and develop AI to find information that is irretrievable using metadata
- › OCR is a prerequisite

FAKTA

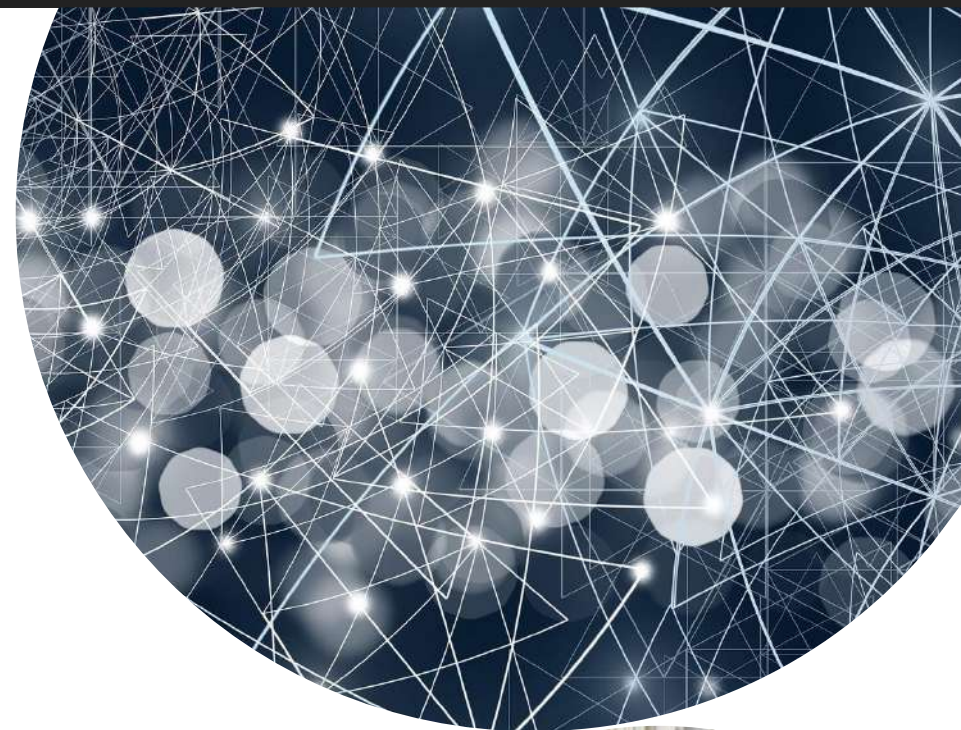


## Anzyz Technologies

- › Norwegian AI
- › Grown from the University in Agder
- › Text analysis: Natural Language Learning (NLP)
- › Language algorithm CCL
- › Skilled in health data

## The pilot dataset and workflow

- › 15 000 patient records from a university hospital
- › Representative for the registry
- › OCR-text is the input to the CCL algorithm
- › AI-base established on random sample of data set
- › Define statistical categories in the Anzyz Dashboard
- › Output is a statistical report (.csv)





# Build statistical categories

## Concept Details

Concept Name

Elins obduksjonsrapport

Phrase Count

10

Target Terms

dødsflekker og dødsstivhet  dødsstivhet  dødsflekker  obduksjonsrapport  
 obdusent  obduksjonsdiagnoser

Ignore Related

Ignore Specific

innholdsfortegnelse obduksjonsrapport  smerter i hoften  
 obduksjonsrapport dødsmelding

Publish

Relevance  High  Low  Negligable

OPEN TERM SELECTION UI

SAVE

DELETE

REPLICATE

# Verify results and gain insight in the data set

Dashboard : Training Demo

Text  
forespørsel

Categories  
in Cities

Mark Dates  Mark Emails  Dynamic Search  Source Please select v

SEARCH  Only List

<p>---U i JUIPostadressePostboks 8014 Dep 0030 OsloBesøksadresseEinar Gerhardsens plass 1 (R4)0030 OsloTlf. 22 33 7010Faks. 22 33 70 12av.tbs IE-post. post@kofa.noNettside: www.kofa.noSak. ... Klagenemndafor offentlige anskaffelserKlagenemndas avgjørelse 26-juni 2003 i sak 2003/108Klager, Topas og Haug Entreprenørforsikring ASinnklagetOslo kommunev/luftsetatenKlagenemndas medlemmer Jens Bugge, Morten Gøler, Siri Teigum.Saken gjelder:Frst for deltagelse i konkurransen. Krav til konkurranseBakgrunn Oslo (heretter kalt "innklagede") kunningjorde i Norsk lysingsblads elektroniskdatabase (DOFFIN) 13.6.2003 en konkurranse med forhandling for grunnarbeid påkyststresbanene på Vestli, Løren og Valle Hovitt. Anskaffelsen ble kurungjort etter reglene forscrift 15.6.2001 om offentlige anskaffelser del III innklagede skode kunningjoring til Norsk lysingsblad 12.5.2003, og det var i denne sakt "Frstfor forespørsel om deltagelse" til 19.5.2003. Topas og Haug Entreprenørforsikring AS (heretter kalt "klager") sendte innklagede et e-postmed forespørsel om å få utlevert konkurransegrunnlaget den26-6-2003 kl. 13.44. Forespørselen ble avvist under henvisning til at fristen for å få utlevertkonkurransegrunnlaget gikk ut 19.5.2003. Parterens anførsler Klager's anførsler:Klager anfører at fristen 19.5.2003 for forespørsel om deltagelse - baje fire virkedager - var så kort at det må anses for å være i strid med regelverket for offentlige anskaffelser. Det vises til klagenemndas avgjørelse i sak 2003/22, som i likhet med den foregående sakgjaldt frst fastsatt av oppdragsgiver for å levere forespørsel om å få delta i konkurransen. Frsteu var i sak 2003/22 fem virkedager. At det der også var tale om å levere inf skatteattestog HMS erklæring kan ikke medføre at vurderingen av kravene til bidrstenes lengde kan bliannerledes.</p>	<p>Creation Time Tue, 01 Jan 2019 01:01:01 GMT</p> <p>File Name 2003108.pdf</p> <p>Page Number 1</p> <p>Total Page Number 3</p> <p>Show Scores</p>
<p>ile illeggSak-: 20.1. av.1: Klagenemndafot offentlige anskaffelserKlagenemndas avgjørelse 26-6-2003 i sak 2003/109Klager, Norske Arkitektors LandsforbundogNorges Prakt sprndc Anktgjordinnklaget. Studentsamskipraden i OsloKlagenemndas medlemmer: Jens Bugge, Morten Gøler, Siri Teigum.Klagesaken gjelder:Forholdet mellom plan- og designkonkurranse og anbudskonkurranse. Tidstriste forinlevering av tilbud. Forhandlingsforbudct ved anbudskonkurranse. Antall deltakere vedbegrenset anbudskonkurranse. Kvalifikasjonskrav og tildelingskriterier. Bakgrunn:Studentsamskipraden i Oslo (heretter kalt "innklagede") kurungjorde a 6.5.2003 i DOFFIN-databasen en konkurranse for anskaffelse av arkitektjenester i kunningjoringen ble bådebetegnelsen "Hastetilfelle - begrenset anbudskonkurranse" og "invitasjon til begrensetarkitektkonkurranse" benyttet. Kontraktens varighet ble angitt å være 18 månedEr og omfattetUtarbeidelse av materiale for forprosjekt, materiale for søknad om rammetilfelle og utarbeidelse av anbudsmateriale for totalentrepris, prosjektering og gjennomføring avovenne". Frst for forespørsel om deltagelse var satt til 6.6.2003. Det var utarbeidet et eget konkurransegrunntag, som ble sendt til leverandørene den19.5/2003. Tilbudsfristen ble oppgitt å være 2.6.2003. Det ble i konkurransegrunntaget stiltfor gende kvalifikasjon skrav i L. Dokumentert erfaring med å studentno i i gprosjekt, z Dokumentert kapasitet til vedlagl fremdriftplan og3 Kontor i Oslo. Tildelingskriterier ene var Pris/honorar Forprosjektet Hele prosjektetPostadressePostboks 8014 Dep 0030 OsloBesøksadresseEinar Gerhardsens plass 1 (R4)0030 OsloTlf. 22337010Faks. 22 33 70 12E-post. post@kofa.noNettside: www.kofa.no</p>	<p>Creation Time Tue, 01 Jan 2019 01:01:01 GMT</p> <p>File Name 2003109.pdf</p> <p>Page Number 1</p> <p>Total Page Number</p>

Autopsy reports

Penicillin allergy

Antibiotic  
resistance

Urinary tract  
infection and  
antibiotic use



# Autopsy reports

## Results of statistical concept autopsy reports

Total records in data set: 15 092

CCL identified as autopsy reports

669

True autopsy reports

86%

False positives

14%

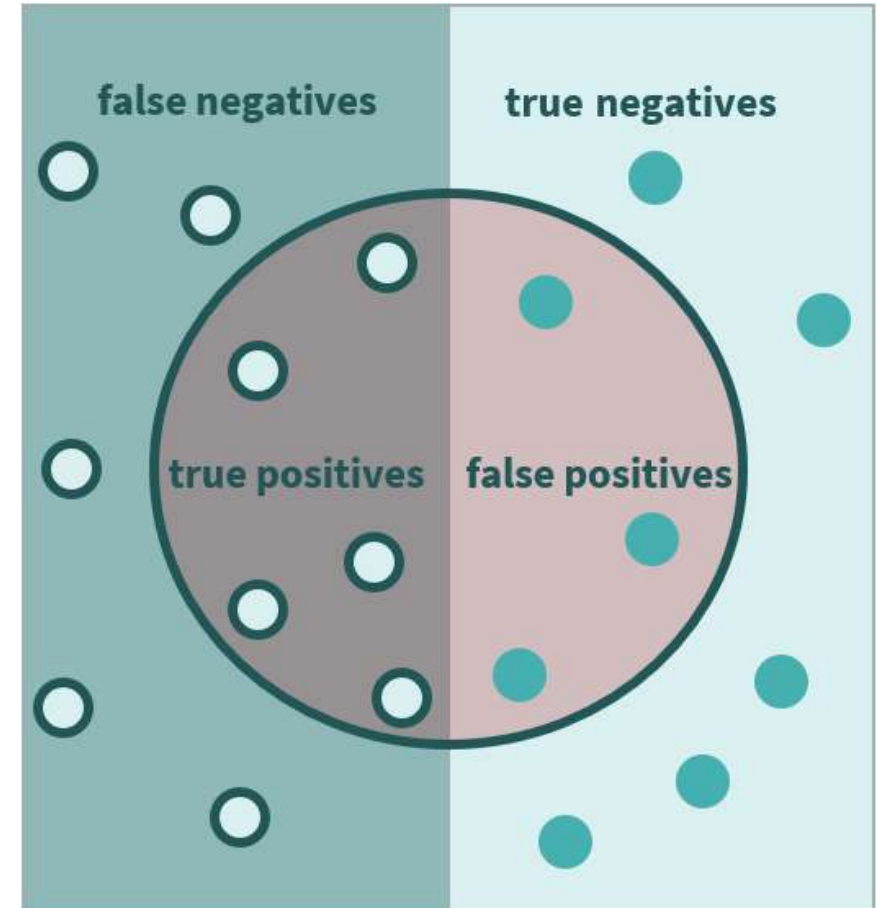
False negatives

0%

True negatives

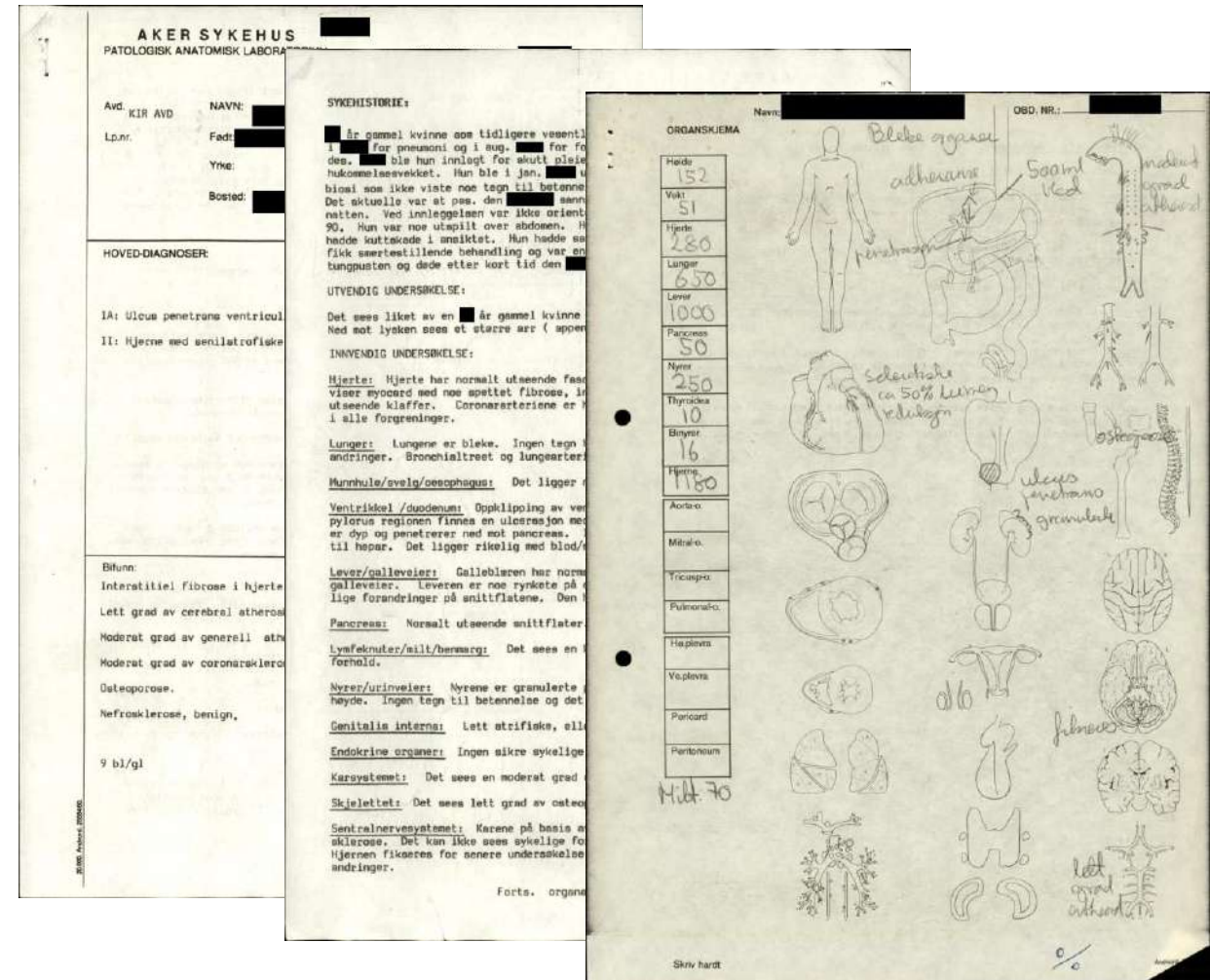
100%

Type	Recall (%)	Precision (%)	F1-measure (%)	F2-measure (%)
PHR which included Autopsy reports	100	86	92	97



# First PhD project hosted by The Norwegian Health Archives Registry

- An interdisciplinary project in three parts that combines:
  - resources and data from NHA
  - modern technology
  - public health knowledge.
- In the first article, AI technology will be further tested in regard to identifying autopsy reports in HARI.





“

*I hope to demonstrate how we can use AI methods to build relevant statistics and gain valuable insight into our extreme amounts of unstructured health data.*

*This way we can offer researchers better and more precise data products.*

---

Elin Bjørnstad-Tuveng, PhD student



Gina Helstad  
The National Archives of Norway  
The Norwegian Health Archives Registry  
[ginhel@arkivverket.no](mailto:ginhel@arkivverket.no)

Kjersti Aasgård  
The National Archives of Norway  
The Norwegian Health Archives Registry  
[kjeaas@arkivverket.no](mailto:kjeaas@arkivverket.no)