

Fra hullkort og strikkepinner til maskinlæring – En reise gjennom Kreftregisterets 70 år med databehandling

Siri Larønningen¹ og Jan F. Nygård²

1) Registeravdelingen, Kreftregisteret, Oslo, Norge

2) Registerinformatikkavdelingen, Kreftregisteret, Oslo, Norge

E-post: sla@krefregisteret.no

SAMMENDRAG

Det har vært en lang reise fra skrivemaskiner, regnemaskiner, hullkort og grafer tegnet på rutepapir til sikre datasentre, skyløsninger, maskinlæring, syntetiske datasett og federerte analyser. Målet med denne artikkelen er å gi et innblikk i behandlingen av data gjennom Kreftregisterets 70 år lange historie, og også si litt om hva vi tror blir viktig i fremtiden. Vi omtaler teknologi for databehandling og deling av data, kodeprosesser, kodeverk og standarder, Kreftregisterets kilder til informasjon, og juss og personvern. Hovedfokuset i artikkelen er databehandling knyttet til Kreftregisterets kjernedata, den såkalte «hoveddatabasen» eller «insidensdatabasen».

Larønningen S, Nygård JF. **From punch cards and knitting needles to machine learning – A journey through the Cancer Registry's 70 years of data processing** *Nor J Epidemiol* 2022; 30 (1-2): 143-154.

ENGLISH SUMMARY

It has been a long journey from typewriters, calculators, punch cards and graphs drawn on grid paper to secure data centers, cloud-based systems, machine learning, synthetic datasets and federated analysis. The goal of this article is to provide some insight to data management through the 70-year history of the Cancer Registry of Norway. We look into technology for data management and sharing of data, coding, classifications and standards, the Cancer Registry sources of information, and laws, regulations and privacy issues. The main focus of the article is data management related to our core data and variables, the so-called “main database” or “incidence database”.

This is an open access article distributed under the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

INNLEDNING

I 1952 hadde Kreftregisteret 3 ansatte og ikke annet kontorteknisk utstyr enn en telefon, et kartotekskap og to mekaniske maskiner; en skrivemaskin og en regnemaskin. Landsforeningen mot Kreft omtaler oppstarten i sitt medlemsblad «Mot kreft» i 1952 (Figur 1).

I dag har Kreftregisteret 208 ansatte [1] og en moderne datapark som blant annet tilbyr elektronisk kommunikasjon, skyløsninger for ikke-følsomme data og et datasenter med optimal sikkerhet for sensitive data. En moderne datapark som hjelper oss med å løse stadig mer komplekse oppgaver. Hva slags reise er det vi har vært med på gjennom disse 70 årene?

ETABLERINGEN AV KREFTREGISTERET

Allerede i 1907 hadde «Det 12te almindelige lægemøde i Kristiania» besluttet å nedsette Den norske komité for kræftforskning, og ut fra mandatet til denne komiteen kom i 1916 Fredrik Georg Gades *Undersøkelser over Kræftsyddommene i Norge* [2].

Under og etter 2. verdenskrig var det økende interesse for og søkelys på forekomst av og årsaker til kreft, blant annet drevet av det presserende behovet for å klargjøre årsaken til voldsomme økningen i lungekreft blant menn i mange vestlige land. Det ble etablert kreftregistre i Upstate New York (1940) og i Connecticut (1941), og i Europa var Danmark først ute med kreftregister fra 1942, under ledelse av den danske kreftpioneren Johannes Clemmensen.

I 1946 ble det holdt en konferanse i København der 12 internasjonale eksperter anbefalte opprettelsen av «Worldwide establishment of cancer registries» for WHO (World Health Organization). Anbefalingen ble oppsummert som følger [3]:

- Great benefit would follow the collection of data about cancer patients from as many different countries as possible
- Such data should be recorded on an agreed plan so as to be comparable
- Each nation should have a central registry to arrange for the recording of such data
- There should be an international body whose duty it should be to correlate the data and statistics obtained in each country

Fire år senere, på WHO's årsmøte i Geneve, beskrev ekspertkomiteen for helsestatistikk hensikten med kreftregistrering på følgende måte:

- Å forbedre mulighetene for at kreftpasientene skulle få den tidligste og beste behandlingen som var tilgjengelig
- Å øke kunnskapene om kreftsykelighet og dens avhengighet av sosiale forhold og miljøfaktorer
- Å gi indikasjoner om vitenskapelige studier rettet mot årsaksforhold ved kreft
- Å gi kunnskaper om overlevelse ved de ulike kreftformer og gi muligheter for kontroll av tiltak iverksatt for å bedre behandlingsforhold

KREFTREGISTERET

I FULL GANG

NYOPPUSSEDE LOKALER ute på Radiumhospitalet er det landsomfattende Krefregister, som er opprettet på initiativ av Landsforeningen mot Kreft, i full virksomhet fra 1. februar i år. I den praktiske utforming av Krefregisterets arbeid er det etablert et nært samarbeid mellom Det Norske Radiumhospitalet, Helsedirektoratet, Landsforeningen mot Kreft samt Statistisk Sentralbyrå.

Den første betingelse for arbeidet er den meldepikt som alle landets sykehus er blitt pålagt av Helsedirektoratet i 1951 og som går ut på at de skal sende inn til Krefregisteret et utfylt skjema over alle kreftpasienter. På dette skjema må de besvare en rekke spørsmål som kartlegger de forskjellige kreftsykdommers forekomst her i landet og de betingelser hvorunder de opptrer.

Tross det arbeid som er utført ute og hjemme av enkelte forskere, institutter og sykehus, er vår viten om kreftens sykdomsformer fremdeles meget mangelfull. Vi vet bare anslagsvis hvor mange nye tilfelle av kreft som oppstår og kommer under behandling her i landet pr. år. Den nåværende offentlige

Fra Krefregisterets kontor (fru Ragnhild Rhodes t. v. og frk. Zaren Wang).



42

norske kreftstatistikk omfatter bare dødsfall. Vi vet lite om årsakene til kreft og om de enkelte kreftformers relasjon til yrke, ernæring, sosiale forhold og miljø i videre forstand. Vår tids medisinske forskning bruker mer og mer statistikkens prinsipper og metoder på medisinske problemer. Ved opprettelsen av Krefregisteret kan disse metoder også bli brukt i kampen mot kreft.

De utfylte meldeskjemaer som kommer inn til Krefregisteret blir gjennomarbeidet og opplysningene blir overført på navne og registerkort. Kortene følger pasientene fra diagnosen er stillet og så lenge de lever, gjennom behandlingen, og den regelmessige kontroll. Alle opplysninger som etter hvert kommer inn om pasienten blir ført inn på kortet. Ved det samarbeid som er etablert med Statistisk Sentralbyrå, er Krefregisteret sikret melding om alle dødsfall blant kreftsyke.

I samarbeid med Statistisk Sentralbyrå vil det også hvert år bli foretatt en bearbeidelse og analyse av det materiale som er tilveiebrakt, og resultatene vil bli publisert.

Det sier seg selv at dette materiale vil gi verdifulle opplysninger om kreftsykdommens geografiske utbredelse. Dessuten vil de gi et riktig utgangsmateriale for studier over spesielle kreftproblemer.

Landsforeningen mot Kreft har gjort en stor innsats både ved planlegging og ved de økonomiske forpliktelser som Landsforeningen har påtatt seg i forbindelse med Krefregisterets opprettelse og drift. Det er Landsforeningen som vesentlig finansierer Krefregisteret. Dr. Einar Pedersen er ansatt som bestyrer av registeret.

Det hersker stor travelhet på Krefregisterets kontor da vi avlegger det et besøk. Innkomne skjemaer blir gransket og opplysningene overført på kort. Oppfordringen til å melde alle kreftpasienter har møtt stor forståelse, og fra sykehus og medisinske institutter sendes nøyaktig utfylte skjemaer — et viktig ledd i kampen mot kreften.

101 SPØRSMÅL

må på grunn av plassmangel utstå til neste nummer.

MOT KREFT

AV INNHOLDET: «Det nytter» av direktør dr. med. R. Eker. «Lungekreft» av dr. med. Leif Efskind. «Kreftdagen 15. okt. 1952». «Krefregisteret i full gang». « $\frac{1}{4}$ mill. kroner til kreftforskning».

NR. 3 — 1952.

Figur 1. Omtale av Krefregisterets oppstart i Landforeningen mot Kreft sitt medlemsblad "Mot Kreft", nr. 3, 1952.

Med bakgrunn i dette foreslo Landsforeningen mot Kreft i 1951 å opprette et nasjonalt krefregister i Norge. Krefregisteret ble etablert i samarbeid med Helsedirektoratet, Radiumhospitalet og Statistisk Sentralbyrå, og med lovhjemmel fra Sosialdepartementet fra januar 1952 [4].

KREFTREGISTERETS INFORMASJONSKILDER

Allerede fra etableringen av Krefregisteret ble det innført en lovhjemmel som påla helsepersonell å rapportere informasjon om kreftsykdom til Krefregisteret, og at pasientene ikke kan reservere seg mot — og heller ikke

Fra Helsedirektøren.

Til oppslag.

Melding til kreftregisteret.

I henhold til Sosialdepartementets bestemmelse av 17. oktober 1951 skal alle tilfelle av kreftsykdommer og visse sykdommer som kan lede til kreft, meldes på fastsatt skjema til Kreftregisteret. Følgende svulstformer og tilstander omfattes av bestemmelsen:

Karsinomer	Maligne nævussellesvulster
Sarkomer	(Maligne melanomer)
Maligne, embryonale nyresvulster	Kraniofaryngiomer
(Wilms' tumor, adenomyosarkom, nefroblastom)	Adamantinomer (Ameloblastomer)
Teratomer, maligne eller benigne	Semimaligne former av spyttkjertelsvulster
Granulosasellesvulster	Papillomer i urinveiene
(med thecomer)	«Polypper» i ventrikkel og colon
Arrhenoblastomer	Chordomer
Seminomer (Dysgerminomer)	Synovialomer
Chorioepitelomer	Conjunctivomer
Karsinoider	Myelomer (plasmocytom etc.)
Insulomer	Lymfogranulomatose
Nevroblastomer	Leukemier
Maligne nevrinomer	Mycosis fungoides
Alle svulster i sentralnervesystemet (også de som er maligne helt eller vesentlig på grunn av sin lokalisasjon)	

Videre skal meldes tilfelle hvor det foreligger begrunnet mistanke om kreft, og tilfelle som hyppig eller alltid er prekankrøse, herunder:

Pagets sykdom eller Pagetoide forandringer i hud	Atypiske epitelhyperplasier
Bowens sykdom	Atypiske, follikulære hyperplasier i lymfoide vev
Lentigo maligna	

VENNLIGST VÆR OPPMERKSOM PÅ DE REGLER SOM GJELDER FOR RAPPORTERINGEN. VIS NØYAKTIGHET VED UTFYLING AV MELDESKJEMAENE.
SVAR OM MULIG PÅ ALLE SPØRSMÅL.

Figur 2. Oppslag om meldeplikt til Kreftregisteret.

trenger å samtykke til – registrering i Kreftregisteret (Figur 2). Denne meldeplikten og hjemmelen til registrering av informasjon ble opprettholdt også da *Lov om personregistre* ble innført i 1978 og Datatilsynet ble opprettet i 1979. Kreftregisteret søkte om og fikk konsesjon i henhold til *Lov om personregistre* i 1982, og fikk en egen forskrift i 2001.

Lovpålagt rapportering fra helsetjenesten, uten hensyn til pasientenes samtykke og uten at pasientene kan be om at informasjonen slettes, er en av de store suksessfaktorene for sentrale helseregistre i Norge. Det sikrer komplette og oppdaterte helseregistre som er gull verdt for både helsetjenesten, forvaltningen, forskerne og, ikke minst, dem de skal betjene – pasientene selv.

Meldinger fra patologer og klinikere i helsetjenesten har i alle år dannet ryggraden av informasjon i Kreftregisteret. Det ble tidlig utarbeidet egne meldeskjema for rapportering til Kreftregisteret – både for patologer (Figur 3) og for klinikere (Figur 4).

Meldeskjemaene skulle fylles ut på papir og sendes til Kreftregisteret per post. Avsenderne ble i noen grad kompensert for rapporteringen, og helt frem til 2016 betalte Kreftregisteret 2,50 kroner per patologisvar fra patologiavdelingene. Kompensasjonen skulle dekke utgifter til toner, papir og frankering av sendingene.

Meldeskjema for patologi ble på begynnelsen av 1970-tallet tatt ut av bruk og erstattet av at Kreftregisteret mottok kopi av selve svarrapporten («patologiremissen») patologene sendte til klinikken. I dag sendes kopi av alle relevante svarrapporter fra patologiavdelingene til Kreftregisteret som elektroniske meldinger på xml-format [5].

Klinikerne har gjennom hele Kreftregisterets 70 år lange historie måttet fylle ut og sende inn meldeskjema. Meldeskjemaene er i dag mer spesifiserte enn tidligere, og er tilpasset både kreftform og hvor pasienten er i utrednings- og behandlingsforløpet. Tilpasningene har sammenheng med fremveksten av kvalitetsregistre på

Figur 3. Tre versjoner av meldeskjema fra patologiavdelingene.

Figur 4. Tre versjoner av meldeskjema fra klinikere.

kreftområdet, der det er krav til mer spesifikk og detaljert informasjon ut over den kjerneinformasjonen Kreftregisteret har registrert siden 1952. Rapporteringen skjer elektronisk via en meldetjeneste utviklet av Kreftregisteret. Figur 5 viser et lite utdrag fra et elektronisk meldeskjema for utredning av lungekreft.

Totalt mottar Kreftregisteret nå over 300 000 patologisvar årlig, og antallet øker for hvert år – både med økende antall krefttilfeller og økende overlevelse, og fordi det utføres flere typer undersøkelser av vevsmaterialet, blant annet utvidede molekylærpatologiske og genetiske undersøkelser. I 2021 mottok Kreftregisteret også i overkant av 86 000 kliniske meldinger. Dette dekker både utrednings- og behandlingsmeldinger for krefttilfeller diagnostisert i 2021, og klinisk informasjon for utredning og behandling som har foregått

tidligere, men hvor Kreftregisteret har måttet etterspørre informasjonen.

En annen viktig kilde til informasjon i Kreftregisteret er dødsattester. Allerede i 1952 ble det innledet et samarbeid mellom Kreftregisteret og Statistisk Sentralbyrå (SSB) for gjensidig kvalitetssikring og komplettering av informasjon [6]. Dødsattestene inneholder informasjon om årsaken eller sekvensen av årsaker som fører til dødsfallet. Her skal både den umiddelbare dødsårsaken (for eksempel “blødning i luftveiene”), og det som forårsaket den umiddelbare dødsårsaken (for eksempel “metastase i lunge” som igjen kom av “brystkreft”) angis. Videre skal sykdommen eller skaden som startet sekvensen som forårsaket død registreres som “underliggende dødsårsak” [7]. Kreftregisteret bruker informasjonen på dødsattestene til å registrere tilfeller som

UTREDNING
Meldes etter avsluttet utredning, før igangsatt behandling

Pasient/behandlingsinstitusjon

Fødselsnummer
01010000000 Ikke norsk personnummer

Navn
Kari Nordmann

Sykehus*
Velg...

Avdeling* ?
Velg...

Funn i utredningen

Primærtumor (med eller uten metastase(r)) *Utredning av primærtumor og samtidig utredning av regionale lymfeknutemetastaser/fjernmetastaser før igangsetting av primærbehandling*

Kun metastase(r) *Dette alternativet skal brukes når man ikke har andre opplysninger om pasienten enn funn og utredning av en metastase fra lungekreft. Skal benyttes der diagnosen lungekreft er stilt på bakgrunn av funn av metastase og man ikke har forsøkt å utrede primærtumor (TX)*

Lokalisasjon primærtumor

Primærtumors lokalisasjon
 Lunge (C34) Trakea (C33)

Har pasienten synkron primærtumor? ?
 Ja Nei

Side*
 Høyre Venstre Ukjent

Spesifiser lokalisasjon*
Velg...

Klinisk sikker kreft* ?
 Ja Usikker

Start utredning

Dato mottatt henvisning til lungelege/spesialisthelsetjenesten (dd.mm.åååå)* ?
 Ukjent

Dato for første innleggelse/konsultasjon (dd.mm.åååå)* ?

Hovedårsak til utredning* ?
 Tumorrelaterte symptomer Tilfeldig funn Ukjent

Funksjonsstatus, ECOG (ved tidspunkt for behandlingsbeslutning)*
Velg...

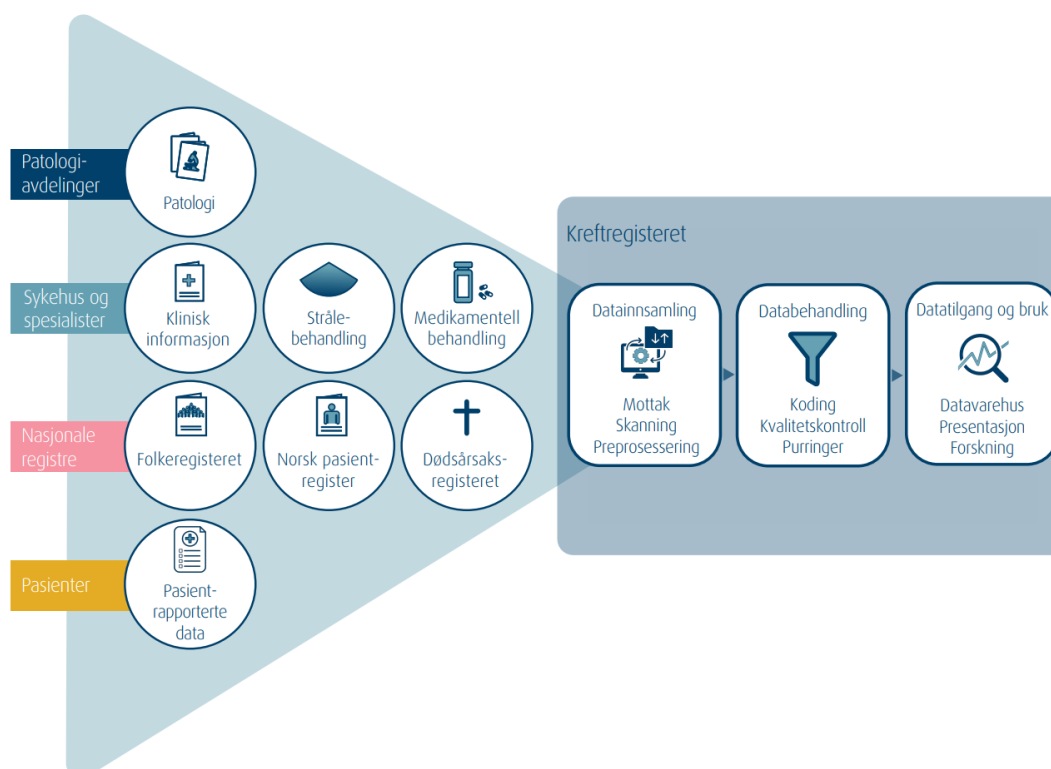
Figur 5. Utdrag fra et elektronisk meldeskjema for utredning av lungekreft.

ikke allerede finnes meldt til eller registrert i Kreftregisteret. I tillegg legger de underliggende dødsårsakene grunnlaget for dødelighetsstatistikken (mortalitetsstatistikken) i Dødsårsaksregisteret. Informasjon om antallet personer som dør av ulike kreftsykdommer over tid er en av grunnpillarene i kreftstatistikken. Kreftregisteret deler også rutinemessig informasjon om kreftpasienter med Dødsårsaksregisteret for kvalitetssikring.

De senere årene har det kommet enda flere kilder til informasjon i Kreftregisteret. I 1997 ble det etablert rapportering fra strålemaskinene ved alle landets stråleenheter, og i 2002 startet rapporteringen av pasientadministrative data fra spesialisthelsetjenesten (med innhenting av data tilbake til 1998). Denne informasjonen ble i utgangspunktet rapportert fra hvert enkelt sykehus, men da Norsk Pasientregister (NPR) i 2010 ble gjort personidentifiserbart, fikk Kreftregisteret hjemmel til å hente relevant informasjon direkte fra NPR heller enn fra hvert enkelt sykehus. I tillegg til disse viktige endringene, ble pasientenes stemme for første gang inkludert som kilde i Kreftregisteret i 2019 gjennom befolkningsbaserte undersøkelser om livskvalitet [8]. I 2020 etablerte Kreftregisteret også et stort samarbeid mellom offentlig, privat og frivillig sektor for å rutinemessig hente inn detaljert informasjon om medikamentell behandling direkte fra sykehusenes systemer [9],

noe som er ekstra viktig i en tid da stadig flere nye medisiner viser seg å kunne ha revolusjonerende effekt for enkelte pasienter, men også kan ha store bivirkninger.

I tillegg til informasjon fra helsetjenesten, mottar Kreftregisteret også jevnlig oppdateringer fra Folkeregisteret. Her ligger et annet av suksesskriteriene til de norske helseregistrene: Norge har god oversikt over befolkningen, takket være eksistensen av fødselsnummer som unikt identifiserer hver enkelt person og gjør det mulig å knytte informasjon fra flere kilder sammen. Fødselsnummer ble innført i Norge i 1964 [10]. Før den tid ble det brukt ulik informasjon til å identifisere enkeltpersoner, som for eksempel fødested, fødselsdato, navn og adresse. Det var imidlertid vanlig at kvinner skiftet navn ved ekteskap, og menn kunne etter eget ønske ta gårdsnavn. Folk flyttet også og endret dermed adresse. Feil eller mangler i navne- og adresseendringer kunne føre til dobbeltregistreringer eller feil i registrene. Innføringen av fødselsnummer gjorde at identifiseringsarbeid og kobling av data både ble enklere og mer treffsikkert. Det gikk imidlertid mange år – helt til 1981 – før sykehusene ble bedt om å påføre personnummer på meldingene i tillegg til fødselsdato, slik at Kreftregisteret fikk det komplette fødselsnummeret på hver enkelt melding. Det lettet



Figur 6. Kreftregisterets kilder til informasjon og videre dataflyt.

arbeidet i Kreftregisteret betraktelig. Alle personer som var i live i 1964, eller er innvandret eller født senere, og har en kreftdiagnose registrert i Kreftregisteret, har i dag fødselsnummer i tillegg til et Kreftregister-spesifikt personløpenummer som unik identifikator. Personer som er registrert i Kreftregisteret og som døde før 1964 har kun det Kreftregister-spesifikke personløpenummeret som unik identifikator. Disse pasientene kan inkluderes i forskning og statistikk med informasjon kun fra Kreftregisteret, men kan ikke enkelt kobles mot andre registre, ettersom de mangler fødselsnummer.

Kreftregisterets kilder per 2022 er vist til venstre i Figur 6. Det at Kreftregisteret har mange uavhengige informasjonskilder, gjør at komplettheten og kvaliteten på registerets data er svært høy [11].

DATABEHANDLING I KREFTREGISTERET

Basert på kunnskap fra kartleggingsarbeidet utført av Fredrik Georg Gade [2] og erfaring fra det danske kreftregisteret kunne man forutsi at Kreftregisteret ville motta tusenvis av rapporter om kreft hvert år. Likevel var det neppe noen som hadde forutsett hvor omfattende og ressurskrevende en slik landsomfattende kreftregistrering skulle bli. I starten var det verken utarbeidet passende kodeverk eller etablert gode rutiner og systemer for databehandling.

Kodeverk

De grunnleggende kodeverkene for registrering av krefttype (morfologi) og sykdommens utgangspunkt (topografi), diagnostiske verktøy (basis for diagnose), kreftens utbredelse (metastase og TNM) og behand-

lingsformer (f.eks. kirurgi) kom på plass i løpet av de første 2-3 årene. De ble utarbeidet av Kreftregisteret i samarbeid med internasjonale grupper (jf. anbefalingene fra WHO 1946). Kreftregisteret fulgte blant annet anbefalingene i WHO handbook for standardized cancer registries [12]. Den internasjonale kreftunionen (UICC) hadde også utviklet et kodeverk som Kreftregisteret tok i bruk. Etter hvert som kreftdiagnostikken og -behandlingen utviklet seg, ble det behov både for å oppdatere og videreutvikle eksisterende kodeverk og for å etablere nye. I dag benytter Kreftregisteret internasjonale kodeverk med norske oversettelser og tilpassninger (for eksempel ICD-O-3 og ICD-10) og nasjonale kodeverk (for eksempel Norpat) der det er relevant og hensiktsmessig, og samarbeider med institusjoner som WHO/IARC, ENCR (European Network of Cancer Registries) og Direktoratet for e-helse for å holde kodeverk og standarder oppdaterte, relevante og så like som mulig for lettere kunne sammenstille statistikk fra ulike institusjoner og mellom land. Kreftregisteret har også etablert egne kodeverk for en rekke variabler der et passende kodeverk ikke har vært tilgjengelig eller hensiktsmessig å benytte. En komplett oversikt over Kreftregisterets variabler og tilhørende kodeverk finnes i dag i Kreftregisterets metadatabase – ELVIS [13].

Databehandling

Databehandlingen i Kreftregisteret har gjennomgått store endringer gjennom disse 70 årene, både med hensyn til hvem som har hatt ansvaret for mottak og behandling av informasjon og hvilke verktøy som har vært i bruk.

A B 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29											
Kreftregisteret		Adr.:			St.:		Lok.:			37-40	
F.:		Yrke:			Met.:		Hist.u.s.:			42	
Kliniske sympt.:		Første sympt.:			Død: 58-61		Spes.u.s.:			52-55	
Søkte lege:		a			I b		Seksjon:			34-5	
Diagnose stilt:		c			II		Op.:				
Behandling startet:		1			2		3			4	
Fødsler		Aborter		Menopause		1-3		44-51		5	
6		7		8		9		10			

Figur 7. Hullkort benyttet i Kreftregisteret.

Koding og registrering

I de første 30 årene av Kreftregisterets historie var det leger som tolket og kodet den medisinske informasjonen før den ble registrert. Et innblikk i dette arbeidet kan leses i artikkelen av Eiliv Lund i dette nummeret [14]. Fra begynnelsen av 1980-tallet ble det ansatt personell med ulik faglig bakgrunn som fikk opplæring i medisinsk koding. I dag er det 25-30 medisinske kodere i Kreftregisteret som jobber med koding og registrering av medisinsk informasjon. Det grunnleggende i dette arbeidet er det samme i 2022 som det var i 1980 og i 1950: å gjøre om tekstlig, beskrivende informasjon til koder i henhold til standard kodeverk, slik at det er mulig å bruke informasjonen i statistikk og forskning. Dette arbeidet er nødvendig fordi mye av informasjonen som kommer til Kreftregisteret, kun finnes som tekstlige beskrivelser. Fra 1980-tallet frem til rundt 2008 var koderne «all-roundere», det vil si at alle koderne jobbet med alle kreftformer. Med fremveksten av kvalitetsregistre og mer spesialisert informasjon om hver kreftform ble også koderne mer spesialiserte og har i dag spesialkompetanse innenfor ulike grupper av kreftdiagnoser. Andre viktige oppgaver er å kvalitetssikre informasjon fra ulike kilder opp mot hverandre, avklare hva som er korrekt informasjon hvis ulike kilder ikke stemmer overens og etterspørre manglende informasjon. Til hjelp i prosessen har koderne blant annet en generell kodebok og mulighet for å rådføre seg med internt eller eksternt medisinsk personell. En del av kvalitetskontrollen er også automatisert gjennom en regelmotor som sjekker og godkjenner om informasjonen er gyldig i henhold til gjeldende kode- og regelverk [15].

Det kreves mye opplæring og erfaring å bli en god koder, og av og til trengs kunnskap som går ut over det rent medisinske, som den gangen vi mottok en beskrivelse av retningen på reseksjonsrendene for et melanom

som «styrbord», «babord», «front», «akter» og «kjøl» i stedet for med mer spesifisert medisinsk terminologi. Selv om dette er et unntak, er det beskrivende for det Kreftregisterets kodere møtes av hver dag: det benyttes ulike terminologier og beskrivelsene finnes på en blanding av norsk, engelsk, svensk, dansk, gresk og latin.

Verktøy for databehandling

Når det gjelder verktøy for databehandling har Kreftregisteret, som de fleste andre deler av samfunnet, vært gjennom en revolusjonerende utvikling de siste 70 årene. Ferden har gått fra hullkort og elektroniske regnemaskiner til sikre datasentre og skyløsninger.

Regnemaskiner har eksistert helt siden kulerammens/abakusens tid. Den første mekaniske kalkulatoren som kunne utføre addisjon og subtraksjon ble bygget i 1650 av Blaise Pascal. Den ble forbedret av tyske von Leibniz i 1673 slik at den også kunne utføre multiplikasjoner [16]. Elektroniske regnemaskiner ble i 1940-årene betraktet som «elektronhjerner» med en voldsom kapasitet til å løse matematiske oppgaver, men påliteligheten var ikke alltid like god. Det ble sagt at en datamaskin kunne gjøre flere regnefeil på ett minutt enn et menneske kunne prestere i hele sitt liv [17]. Det regnemaskinene også manglet var en mulighet for å kunne lagre resultatene for videre behandling. Hullkortene ble løsningen på dette, og var fra 1920-årene til langt ut i 1960-årene det mest brukte lagringsmedium for data.

De første hullkortene som ble benyttet i Kreftregisteret, randhullkortene – også kalt «storkort» – kunne ikke bearbeides maskinelt. Storkortene inneholdt et tekstlig utdrag av den medisinske informasjonen som ble sendt til Kreftregisteret. Legene kodet deretter informasjonen direkte på storkortene og kortene ble klassifisert ved hjelp av hull som ble klippet i kanten – omtrent slik man klippet orienteringsposter eller togbilletter tidligere. Som eksempel kunne et klippet hull bety «kvinne» og

den samme posisjonen uklippet bety «mann». Når det skulle gjøres opptellinger, ble en «strikkepinne» stukket gjennom hullene. Den gjorde at man kunne skille kort som var klippet fra kort som ikke var det for den aktuelle informasjonen. Deretter ble regnemaskinen brukt til å for eksempel beregne insidensrater. Et eksempel på et randhullkort ses i Figur 7.

Kreftregisteret gikk etter hvert over fra de ikke-maskinelt håndterbare storkortene til maskinlesbare hullkort. Det krevdes spesialutstyr både for å lage og lese av disse hullkortene. Det fantes flere spesialiserte hullkortmaskiner [17]:

- *Sorterer* for sortering og telling
- *Kollator* for sammenligning og kontroll
- *Tabulator* for summering og utskrift
- *Kalkulator* for beregninger
- *Reproduser* for kopiering og omredigering av hullkort
- *Interpreterer* for oversetting av hullkoden til tekst
- *Punchemaskin* for å fremstille hullkortene
- *Kontrollpunch* for å kontrollere at hullkortet var riktig punchet

Mot slutten av 1950-tallet kjøpte Kreftregisteret inn en punchemaskin og en kontrollpunch (verifier), samt en kombinert sorterer og tabulator. Bakgrunnen var et stort internasjonalt kreftprosjekt med National Cancer Institute (NCI) i USA, Frankrike og England som ønsket å ha med norske kreftdata i prosjektet etter å ha sett på datakvaliteten i Kreftregisteret. Betingelsen var at Kreftregisteret måtte få datateknisk utstyr. Landsforeningen mot Kreft (nå Kreftforeningen) stilte velvillig inntil 150 000 kroner til disposisjon for innkjøp av disse maskinene – det utgjør i overkant av 2 millioner kroner i 2022-verdi – og fikk siden utgiftene for kjøp av utstyret refundert fra NCI.

Allerede i 1961 skaffet Kreftregisteret mer datateknisk utstyr: en «automatisk reproduserende hullemaskin» og en «kortvelger». Førstnevnte maskin ble brukt til å kopiere opplysninger fra gamle til nye hullkort, mens kortvelgeren kunne plukke ut kort med bestemte kjennetegn og kontrollere rekkefølgen i et hullkortsett.

Hullkortene hang med helt frem til 70-tallet. En av de viktigste begrensningene for hullkortene, var at de ikke var egnet for store datamengder og at de satte begrensninger for hvor mye data som kunne lagres per kort. De vanligste hullkortene var på 80 bytes/karakterer, en ganske beskjeden lagringskapasitet. Hvis Kreftregisteret hadde beholdt hullkortene og lagret 80 bytes per krefttilfelle, ville alle krefttilfeller rapportert i Cancer in Norway for perioden 1953-2021 ($n = 1,3$ millioner) krevd rett i overkant av 100 MB. Den beskjedne lagringskapasiteten var en utfordring for de fleste registre [18], og var en viktig årsak til at utviklingen av bedre dataverktøy og mer lagringsplass satte fart internasjonalt [16].

Med den datatekniske utviklingen endret også arbeidsmetodene i Kreftregisteret seg. I 1969 ble hele Kreftregisterets historiske materiale (ca. 150 000 kreft-

pasienter) lagt inn på magnetbånd ved Norsk Regnesentral. På midten av 1970-tallet gjorde mikrofilm/mikrofiche-maskinen sitt inntog i Kreftregisteret, og alle meldingene ble fra 1975 mikrofilmet. I ettertid er også de gamle storkortene og hullkortene fra 1953 til 1974 lagt over på mikrofilm. Kreftregisteret har fortsatt en mikrofilm-maskin for avlesing og utskrift av disse meldingene hvis det skulle være nødvendig med en manuell kontroll av gamle data.

Utover mot slutten av 1970- og starten av 1980-tallet gjorde terminalene og en sentral server sitt inntog. Punchemaskinene forsvant i denne perioden ut av bruk, og lagringen av data skjedde direkte på server og på eksterne lagringsmedier, som for eksempel magnetbånd. Parallelt med utviklingen i resten av verden gikk Kreftregisteret mot slutten av 80-tallet over til PC'er, det vil si datamaskiner for hver enkelt bruker, som ble knyttet sammen med hverandre og med servere i et nettverk. Skannere ble også tatt i bruk, og Kreftregisteret gikk over fra å mikrofilme hver enkelt melding til å skanne dem inn og lagre dem som bildefiler. Bildefilene ble så gjort tilgjengelige på riktig pasient i det internt utviklede kreftregistreringsprogrammet og knyttet til registrering av den spesifikke meldingen. Kreftregisterets kodere kunne dermed hente opp og vise originalmeldingen i registreringsprogrammet. En lang stund foregikk koding direkte på papirmeldingen før meldingen ble skannet og registrert inn. Vi måtte noen år inn på 2000-tallet før denne prosessen ble snudd om, slik at meldingene først ble skannet inn og knyttet til rett pasient, før koderne hentet uregistrerte meldinger frem fra en meldingskø og kodet og registrerte informasjonen direkte inn i registreringssystemet. Dette er fortsatt arbeidsprosessen i Kreftregisteret i 2022 (Figur 6). Skanning av papirmeldinger har helt til nå vært en viktig del av databehandlingen. Det var først i juni 2022 at de siste patologiavdelingene gikk over fra å sende papirkopier av patologisvarene til å sende elektroniske meldinger som xml-filer via Norsk Helsenett. Forsatt mottar Kreftregisteret motvillig noen papirmeldinger – tykkeligvis i betraktelig mindre omfang enn tidligere.

På 90-tallet ble det gjennomført et stort arbeid for å overføre mikrofilm-meldinger til separate bildefiler som kunne lagres på server sammen med de skannede meldingene. I 2014, før Kreftregisteret flyttet fra Majorstuen til Montebello, ble også alle de gamle meldingene som lå til grunn for storkortene hentet opp fra Kreftregisterets kjellerarkiv og skannet over til bildefiler. Det gamle papirmaterialet ble deretter makulert, og alle meldingene er bevart i elektronisk form på Kreftregisterets servere.

Med data kodet, registrert og lagret i Kreftregisterets servere lå også verden mer åpen for enklere tilgang til og analyse av data. Borte var elektroniske regnemaskiner og håndtegnede insidensgrafer på ruteark, og inn kom i større og større grad databehandling ved hjelp av PC og spesiallagde databehandlings- og statistikkprogrammer, som SPSS og Stata. Noe av den forskningshistoriske utviklingen i Kreftregisteret er omtalt i

en artikkel av Grimsrud [19]. Med den teknologiske utviklingen kom også en strengere lovgivning rundt bruk av kreftdata til ulike formål. I dag er bruk av data fra Kreftregisteret blant annet regulert i EUs personvernforordning (GDPR) [20] og Norges adaptasjon av denne i Personopplysningsloven [21], i Helseregisterloven [22], i Helseforskningsloven [23] og i Kreftregisterforskriften [24].

FREMtiden FOR KREFTREGISTERET

Den teknologiske utviklingen de siste ti årene har vært eksplosiv, og vi befinner oss i en verden der mange går rundt med mobiltelefoner med mye større datakraft enn det som i det hele tatt var mulig i Kreftregisterets ungdomstid – eller for bare 20 år siden. Samtidig henger vi etter på noen områder. Som nevnt mottok Kreftregisteret fortsatt i juni 2022 papirmeldinger, klinikerne må manuelt logge seg inn i KREMT-portalen for å sende oss informasjon som allerede ligger dokumentert i det elektroniske pasientjournalssystemet (EPJ) og vi trenger fortsatt 25-30 kodere som kan gjøre om beskrivende tekst til koder og foreta datakvalitetskontroll. Hvilke retninger og utviklingstrekk bør Kreftregisteret følge fremover for å sikre fortsatt effektiv, eller enda mer effektiv, kreftregistrering med større datamengder og fortsatt høy kvalitet?

Rapportering, koding og registrering

En forutsetning for effektiv databehandling er datastandarder, og innsending av data på en standardisert måte. Dessverre kan standarder være fortolket ulikt og variere mellom mottakere og avsendere. Oppfølging og vedlikehold av standarder krever store ressurser og mye tid, men hvis dette viktige arbeidet ikke har fokus og ressurser kan det føre til inkonsistent informasjon internt i registre og på tvers av ulike registre. Et hovedsatsningsområde for Kreftregisteret nå og fremover vil være å bidra til standardisering og strukturering av data i originalkildene for økt gjenbruk på tvers av ulike kilder. Det vi ønsker, er at den viktigste delen av informasjonen i EPJ-systemene og i patologenes LIMS-systemer (Laboratory Information Management System) skal være lagret på en måte som gjør det mulig å automatisk hente ut relevant informasjon og gjenbruke den i Kreftregisterets systemer – og andre steder – uten at en kliniker eller patolog manuelt må gjøre en ekstra jobb, og uten at Kreftregisterets kodere må tolke informasjonen og registrere den inn på nytt. På den måten legger vi til rette for automatisk koding av informasjon, basert på strukturert og standardisert informasjon direkte fra helsetjenesten.

Det er flere prosesser og prosjekter på gang for å nærme seg denne type løsninger. Begge de store EPJ-systemene i Norge – DIPS Arena og Helseplattformen (EPIC) – jobber med å strukturere relevant klinisk informasjon ved bruk av f.eks. arketyper og SNOMED CT slik at denne informasjonen kan deles med andre systemer. Tilsvarende foregår det et arbeid med å få på plass en nasjonal modul for synoptiske svrappporter for patologi, dvs. diagnose- og prøvespesifikke strukturerte

oppsummeringer av de viktigste funnene i en patologirapport. Kreftregisteret er involvert i disse prosjektene blant annet gjennom samarbeid med Direktoratet for e-helse, DIPS og Helse Midt for å få til strukturert klinisk rapportering fra EPJ, og med Helse Vest og prosjekt Nasjonal digital patologi for en nasjonal modul der patologiinformasjon fra vevsprøver tatt i forbindelse med tarmscreeningprogrammet er tenkt som pilotprosjekt. Parallelt pågår det også prosjekter i Helse Vest og i Helse Sør-Øst for å bruke robotstyrte prosesser (Robotic Process Automation – RPA) til å hente ut informasjon fra EPJ og fylle dette inn i relevant meldeskjema i KREMT-portalen. Helse Vest har kommet lengst i dette arbeidet, og rapporterer rutinemessig klinisk informasjon om prostatakreft ved hjelp av en RPA-prosess. Kreftregisteret har i tillegg gjennomført flere prosjekter med bruk av språkbehandling, såkalt Natural Language Processing (NLP) og maskinlæringsalgoritmer, for å se om det kan erstatte den jobben koderen gjør med å lese og kode tekst i patologisvar [25-27]. Dette fungerer greit i mindre skala, for én kreftform eller en bestemt type prøve, men vi ser utfordringer med å skalere dette opp til noe som kan fungere for de fleste patologisvarene Kreftregisteret mottar. Spesielt er dette utfordrende fordi patologisvarene per i dag, som tidligere nevnt, er ulike både i form og i språk.

Et spesielt spennende felt fremover er bruk av ny teknologi i diagnostikk og behandling av kreftpasienter. Kreftregisteret er også involvert i flere prosjekter rundt dette tema. Ett eksempel er knyttet til spørsmålet om kunstig intelligens kan hjelpe røntgenlegene i Mammo-grafiprogrammet med å bli enda mer presise og effektive i sin leting etter brystkreft hos kvinner [28]. Et annet eksempel er om maskinlæring kan være til hjelp i overgangen fra en generell, standardisert kreftscreening for livmorhalskreft til en mer personilpasset screening basert på hver enkelt person sin individuelle risiko for kreftutvikling [29,30].

Databehandling

Kreftregisteret har en lang tradisjon for å selv utvikle den programvare som er nødvendig for god registerdrift, og har en egen Registerinformatikkavdeling som har både drift- og nettverkskompetanse, sikkerhet- og arkitekturkompetanse, utviklere og forskere innen registerinformatikk. Et nært og godt samarbeid mellom Registerinformatikkavdelingen og de øvrige avdelingene i Kreftregisteret er essensielt for drift og videre utvikling av nødvendige systemer og applikasjoner.

De siste fem årene har mye av tiden gått til å utvikle og ta i bruk et nytt kode- og registreringsystem som skal erstatte det gamle systemet som ble tatt i bruk på slutten av 1990-tallet. Det nye systemet er bygget på en moderne arkitektur og er et modulbasert prosess-system som understøtter koderegistreringsprosessen, i motsetning til det gamle som i all hovedsak var et rent registreringsystem. Det nye systemet er mer fleksibelt og kan håndtere den økende mengden data som skal behandles i Kreftregisteret. Sentralt i systemet er regelmotoren GURI [15] som blant annet brukes til å

validere informasjonen i ulike meldinger, f.eks. til å sjekke at det kun er brukt gyldige koder og at koder i ulike meldinger for samme krefttilfelle ikke er i konflikt med hverandre, og til å danne (generere) det som er kjerneinformasjonen for et krefttilfelle. Sammen med metadatasystemet ELVIS [13] er Kreftregisteret godt rustet for en fremtid med nye og mer omfattende datakilder, som for eksempel gentypering av kreftsvulster, genomsekvenser, PET, CT og radiologiske bilder.

Deling og analyser av data

Med økt teknologisk utvikling og en generelt større tilgang til informasjon om enkeltpersoner – for eksempel gjennom sosiale media – er det viktig at helseregistre og andre databehandlingsansvarlige er bevisst sin del av ansvaret for å beskytte sensitive data, samtidig som man legger til rette for at data kan brukes til nødvendig kvalitetssikring, forskning og utvikling. Dette er en vanskelig balansegang, og krever samarbeid mellom både helsepersonell, jurister, teknologer og registereksperter. Det er både internasjonalt og her i Norge nå vanlig at det settes opp tekniske løsninger der man sikkert kan oppbevare sensitive data og også har tilgang til datamaskiner med stor kapasitet til å utføre de ulike matematiske oppgavene som kreves for ulike analyser. I tillegg til vårt eget datasenter, benytter Kreftregisteret egne serverer på TSD (Tjenester for sensitive data) [31] som organiseres og driftes av Universitetet i Oslo. Andre eksempler på tilsvarende løsninger i Norge er SAFE (Sikker adgang til forskningsdata og e-infrastruktur) [32] som drives av Universitetet i Bergen og HUNT Cloud [33], som styres fra NTNU.

Kreftregisteret, og andre registermiljøer både i Norge og internasjonalt, har i mange år også jobbet for å få frem gode løsninger for å gjøre statistikk lett tilgjengelig, for eksempel via statistikkbanker på nett. Kreftregisteret har sin egen statistikkbank, STINE [34], med tabeller og kart som viser forekomst av kreft i Norge, og har også sekretariatsfunksjon for NORDCAN [35]. NORDCAN er et samarbeid mellom nordiske kreftregistre og IARC (International Association for Research on Cancer) for å gjøre informasjon om forekomst, dødelighet, prevalens og overlevelse av kreft i de nordiske landene lett tilgjengelig. Denne type statistikk-løsninger kommer til å fortsette å være viktige også i tiden fremover.

Nye teknologier

Tre konkrete områder Kreftregisteret aktivt arbeider med på området deling og bruk av data, er blokkjede-teknologi og smarte kontrakter, syntetiske datasett, og federerte analyser.

Blokkjede eller blokk-kjede, av engelsk «block-chain», er en oversikt over transaksjoner som er arrangert i sammenhengende blokker, og er den teknologien som ligger til grunn for blant annet kryptovaluta [36]. Det er en metode for å utveksle eller dele data på en sikker måte, uten at dataene kan endres underveis og med full sporbarhet i databehandlingen. Blokkjede-

teknologi kan bruke smarte kontrakter, det vil si avtaler som utføres uten at det krever menneskelig interaksjon. Kreftregisteret er blant annet involvert i et prosjekt i samarbeid med Universitetet i Oslo og Simula for å se om denne type teknologi kan brukes til å dele sensitive data mellom ulike institusjoner i helsetjenesten og med helseregistre på en sikker og effektiv måte, og samtidig gi innbyggere/kreftpasienter innsyn og kontroll med egne data. Kreftregisteret ser også på mulighetene for å bruke såkalte consortium blockchains (nettverk det kreves godkjenning for å bli en del av) for å utføre sikre statistiske analyser på helsedata som ligger lagret på flere ulike steder [37].

En av utfordringene i dag er at både forskningsprosjekter og utviklingsprosjekter (for eksempel utvikling av ny teknologi) krever tilgang til data som gjenspeiler virkeligheten. Da har løsningen ofte vært å dele de «ekte» dataene, altså data som er knyttet til faktiske personer. Selv om direkte personidentifiserende informasjon, som navn, fødselsnummer og adresse, er fjernet, er det likevel en risiko for at noen kan klare å identifisere enkeltpersoner i datasettet. Det kan for eksempel skje hvis kombinasjonen av informasjon for én registrering er så unik at den kun passer på én eller noen veldig få personer, eller hvis noen allerede har litt kunnskap om vedkommende og dermed kan finne igjen personen i datasettet, og så tilegne seg mer informasjon enn de i utgangspunktet hadde. Helseregistre jobber svært aktivt med ulike tiltak for å redusere denne risikoen, men med økt mengde av informasjon åpent tilgjengelig blir det også vanskeligere å vurdere risikoen for at kombinasjonen av åpne og sensitive data kan føre til at enkeltpersoner kan identifiseres. En mulig løsning på dette er å utarbeide syntetiske datasett som kan brukes til forskning og til testing av systemer. Syntetisering av datasett er en metode for å lage et datasett som har de samme statistiske egenskapene som et annet datasett, men uten at det finnes informasjon som kan knyttes direkte til enkeltpersoner. Anonymitet og personvern i datasettene beregnes ved bruk av ulike matematiske metoder [38,39]. Bruk av syntetiske datasett blir stadig viktigere i prosjekter i grenseland mellom teknologi-utvikling og kreftforskning, og Kreftregisteret samarbeider med Statistisk sentralbyrå og Sikt-Kunnskapssektorens tjenesteleverandør, og med internasjonale miljøer, for å utvikle og benytte syntetiske datasett.

Deling av reelle data er, som nevnt over, en utfordring. En annen måte å sikre personvernet og minske risiko for identifisering av enkeltpersoner på, er systemer for federerte analyser. Federerte, eller distribuerte, analyser er en arkitektur der de originale dataene blir liggende sikkert i hvert enkelt system og det kun er resultatene av ulike opptelling og analyser som deles.

Ett eksempel på hvordan man tradisjonelt har løst dette, er NORDCAN-prosjektet [35]. Her er det utviklet en applikasjon som kvalitetssikrer, harmoniserer, beriker og analyserer datasettet lokalt i hvert av de nordiske landene. Resultatet fra programmet er anonyme opptelling og analyser som kan sendes fra registre til

IARC og gjøres tilgjengelig i NORDCAN-databasen. Hvert enkelt land må sørge for å lage datasettet som ligger til grunn for statistikken etter gitte kriterier, slik at dataene som kommer ut, er sammenlignbare. Ulempen med denne løsningen, er at det ikke er mulig å slå sammen analyser fra ulike datakilder. Man kan for eksempel ikke se på overlevelse i Norge og Sverige totalt – kun hvert land for seg.

En nyere måte å løse dette på, er å benytte mer avansert teknologi der algoritmene eller analysene sendes elektronisk til hvert enkelt register når det er behov for det. Analysen utføres deretter på hvert datasett i hvert register, og resultatene sendes tilbake og kobles sammen. ”The Personal Health Train” [40] er et eksempel på et slik rammeverk, og vantage6 [41,42] er en teknisk implementasjon av et slik system som det nederlandske kreftregisteret har stått i spissen for. Federerte analyser er blant annet benyttet i en beskrivelse av prognostiske faktorer for overlevelse etter munnhulekreft i Nederland og Taiwan [43].

Bruk av federerte analyser krever standardiserte datasett ute i de ulike datakildene. Federerte analyser må derfor gå hånd i hånd med prosjekter for å lage standardiserte datasett. Ett eksempel på et slikt internasjonalt prosjekt, hvor Kreftregisteret er med, er EHDEN – European Health Data & Evidence Network [44] der datasettene skal mappes til en felles datamodell, OMOP CDM (Observational Medical Outcomes Partnership Common Data Model). Flere og flere prosjekter forsøker å etablere og ta i bruk denne type teknologi i dag, men det kreves fortsatt både juridiske, teknologiske og administrative avklaringer før potensialet kan benyttes fullt ut.

DET VI IKKE HAR REGISTRERT (I DENNE ARTIKKELEN)

Vi har i denne artikkelen i hovedsak konsentrert oss om historikken og fremtidige aktiviteter knyttet til Kreft-

registerets kjernedata, den såkalte «hoveddatabasen» eller «insidensdatabasen». Kreftregisteret er imidlertid mer enn dette. Vi har blant annet ansvaret for tre screeningprogram (Mammografiprogrammet, Livmorhalsprogrammet og Tarmscreeningprogrammet), og drifter også Janus serumbank.

Et eksempel på et stort og tidkrevende prosjekt de siste årene, er utvikling av nødvendige systemer for det nyetablerte Tarmscreeningprogrammet og integrering av disse mot eksisterende systemer og applikasjoner i Kreftregisteret, som for eksempel det nye kode- og registreringssystemet. Modernisering av systemene i Mammografiprogrammet og Livmorhalsprogrammet og integrering av disse mot eksisterende systemer står også på kartet for fremtiden, sammen med en reorganisering av databasen i Janus serumbank slik at vi sikrer gjenbruk av analyser og dermed får en mer effektiv forvaltning av det serum som er lagret fra hver enkelt person. I tillegg kommer nye og spennende prosjekter som for eksempel bruk av kunstig intelligens i mammografiscreening, hjemmetest i livmorhalsprogrammet og genotyping og serum RNA-analyser i Janus.

I Norsk Epidemiologi sitt temanummer om Kreftregisterets 70-årsjubileum finnes gode artikler som går mer i dybden på det vi ikke har dekket i vår artikkel [45-48].

KONKLUSJON

Kreftregisteret har gjennom 70 år samlet inn og opparbeidet seg omfattende erfaring og kunnskap om teknologi, personvern, registerdrift, screeningprogram, biobanker, kreftdiagnostikk, kreftbehandling og kreftforskning. Det er i skjæringspunktet mellom alle disse disiplinene Kreftregisteret har sin styrke og skal fortsette å utvikle seg fremover, slik at vi sikrer rike norske kreftdata av høy kvalitet til bruk for pasientens beste, også i fremtiden.

REFERENCES

1. Kreftregisteret. *Årsrapport 2021*. Oslo: Kreftregisteret, 2021.
2. Gade FG. *Undersøkelser over Kræftsygdommene i Norge*. Kristiania: Videnskapsselskapet, 1916.
3. Wagner G. History of cancer registration. I: Jensen OM, Parkin DM, MacLennan R, Muir CG, Skeet RG, red. *Cancer Registration: Principles and Methods*. IARC Scientific Publications no. 95, 1991.
4. Arnesen K. Opprettelsen av Kreftregisteret. *Tidsskrift for Den norske legeforsking* 2001; **121**: 1256-7.
5. Direktoratet for e-helse. *Svarrapportering av medisinske tjenester v1.4*, 2022. Available from: <https://www.ehelse.no/standardisering/standarder/svarrapportering-av-medisinske-tjenester-v1.4>.
6. Arctander S. *Den offisielle statistikkens historie, Bind IV, Utkast*. Oslo: Statistisk sentralbyrå, 1972.
7. Folkehelseinstituttet. *Slik skal dødsmeldingen fylles ut*, 2021. Available from: <https://www.fhi.no/hn/helseregistre-og-registre/dodsarsaksregisteret/slik-skal-elektronisk-dodsmelding-fylles-ut/>.
8. Kreftregisteret. *Befolkningsundersøkelser om helse og livskvalitet*, 2022. Available from: <https://www.kreftregisteret.no/Registrene/livskvalitet/>.
9. Kreftregisteret. *INSPIRE – informasjon om medikamentell kreftbehandling*. 2022; Available from: <https://www.kreftregisteret.no/Forskning/Prosjekter/inspire/>.
10. Wikipedia. *Fødselsnummer*, 2022. Available from: <https://no.wikipedia.org/wiki/F%C3%B8dselsnummer>.
11. Kreftregisteret. *Cancer in Norway 2021 – Cancer incidence, mortality, survival and prevalence in Norway*. Oslo: Kreftregisteret, 2022.
12. WHO handbook for standardized cancer registries. Geneva: World Health Organization, 1976.

13. Kreftregisteret. *ELVIS – Kreftregisterets metadatabase*, 2022. Available from: <https://metadata.kreftregisteret.no>.
14. Lund E. En sen høstdag ved Kreftregisteret, 1979. *Norsk Epidemiologi* 2022; **30** (1-2): 113-115.
15. Hammad M, et al. IOCL: An interactive tool for specifying, validating and evaluating OCL constraints. *Science of Computer Programming* 2017; **149**: 3-8.
16. Moberg T. *Fra hullkort til gigabytes*, 2019. Available from: <https://www.dvikan.no/hullkort>.
17. Aurbakken E. *Fra hullkort til PC. Glimt fra databehandlingens historie i Statistisk sentralbyrå 1959-1990*. Oslo, Statistisk sentralbyrå, 1999.
18. Irgens LM, Medisinsk fødselsregisters fødsel og senere vekst; noen observasjoner og refleksjoner fra hornet på veggen. *Norsk Epidemiologi* 2017; **27** (1-2): 3-7.
19. Grimsrud TK, Samle, telle, regne, fortolke – tilbakeblikk på Kreftregisterets statistikkproduksjon og forskning gjennom 70 år. *Norsk Epidemiologi* 2022; **30** (1-2): 127-141.
20. Store norske leksikon. *Personvernforordningen*, 2018. Available from: <https://snl.no/Personvernforordningen>.
21. Justis- og beredskapsdepartementet. *Lov om behandling av personopplysninger (personopplysningsloven)*, 2018. Available from: <https://lovdata.no/dokument/NL/lov/2018-06-15-38/>.
22. Helse- og omsorgsdepartementet. *Lov om helseregistre og behandling av helseopplysninger (helseregisterloven)*, 2014. Available from: <https://lovdata.no/dokument/NL/lov/2014-06-20-43>.
23. Helse- og omsorgsdepartementet. *Lov om medisinsk og helsefaglig forskning (helseforskningsloven)*, 2008. Available from: <https://lovdata.no/dokument/NL/lov/2008-06-20-44>.
24. Helse- og omsorgsdepartementet. *Forskrift om innsamling og behandling av helseopplysninger i Kreftregisteret (Kreftregisterforskriften)*, 2001. Available from: <https://lovdata.no/dokument/SF/forskrift/2001-12-21-1477>.
25. Dahl A, Özkan A, Dalianis H. Pathology text mining-on Norwegian prostate cancer reports. I: *2016 IEEE 32nd International Conference on Data Engineering Workshops (ICDEW)*. IEEE, 2016.
26. Weegar R, Dalianis H. Creating a rule based system for text mining of Norwegian breast cancer pathology reports. I: *Proceedings of the Sixth International Workshop on Health Text Mining and Information Analysis*, 2015.
27. Weegar R, Nygård JF, Dalianis H. Efficient encoding of pathology reports using natural language processing. I: *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017*, 2017.
28. Larsen M, et al. Artificial intelligence evaluation of 122 969 mammography examinations from a population-based screening program. *Radiology* 2022; **303** (3): 502-511.
29. Soper BC, et al. A hidden Markov model for population-level cervical cancer screening data. *Stat Med* 2020; **39** (25): 3569-3590.
30. Meng R, et al. Hierarchical continuous-time inhomogeneous hidden Markov model for cancer screening with extensive followup data. *Stat Med Res* 2022; 09622802221122390.
31. Universitetet i Oslo. *Tjenester for Sensitive Data (TSD)*. 2022; Available from: <https://www.uio.no/tjenester/it/forskning/sensitiv/>.
32. Universitetet i Bergen. *SAFE*, 2022. Available from: <https://www.uib.no/safe>
33. Norges teknisk-naturvitenskapelige universitet – NTNU. *HUNT Cloud Documentation*, 2022. Available from: <https://docs.hdc.ntnu.no/>
34. Kreftregisteret. *STINE – Kreftregisterets statistikkbank*, 2022. Available from: <https://sb.kreftregisteret.no/>.
35. Larønningen S, et al., *NORDCAN: Cancer Incidence, Mortality, Prevalence and Survival in the Nordic Countries, Version 9.2*. Association of the Nordic Cancer Registries, Cancer Registry of Norway, 2022.
36. Store norske leksikon. *Blokkjede*, 2022. Available from: <https://snl.no/blokkjede>.
37. Veeraragavan NR, Nygård JF. DeCanSec: A decentralized architecture for secure statistical computations on distributed health registry data. I: *The 16th International Conference on Availability, Reliability and Security*, 2021.
38. Goncalves A, et al. Generation and evaluation of synthetic patient data. *BMC Med Res Methodol* 2020; **20** (1): 1-40.
39. Wagner I, Eckhoff D. Technical privacy metrics: a systematic survey. *ACM Comput Surv* 2018; **51** (3): 1-38.
40. *The Personal Health Train Network*, 2021. Available from: <https://pht.health-ri.nl>.
41. Vantage. vantage6. 2022, Available from: <https://vantage6.ai>.
42. Moncada-Torres A, et al. VANTAGE6: an open source priVAcy preserviNg federaTed leArninG infrastruCTurE for Secure Insight eXchange. I: *AMIA Annual Symposium Proceedings*. American Medical Informatics Association, 2020.
43. Geleijnse G, et al. *Prognostic factors for survival in patients with oral cavity cancer: a comparison of the Netherlands and Taiwan using privacy-preserving federated analyses*. Available at SSRN 3539690, 2020.
44. European Health Data & Evidence Network (EHDEN), *EHDEN: Becoming the trusted open science community built with standardised health data via a European federated network*, 2022.
45. Holen ÅS, Mangerud G, Hofvind S. Mammografiprogrammet – tidligere, i dag og i fremtiden. *Norsk Epidemiologi* 2022; **30** (1-2): 61-68.
46. Bjørge T, et al. CervicalScreen Norway – A screening programme in transition. *Norsk Epidemiologi* 2022; **30** (1-2): 55-60.
47. Hoff G, et al., Norway – a retarded country close to 20 years since EU recommended colorectal cancer screening? A failure or success? *Norsk Epidemiologi* 2022; **30** (1-2): 83-86.
48. Langseth H, et al. Janus serumbank – jakten på biomarkører for kreft. *Norsk Epidemiologi* 2022; **30** (1-2): 87-92.