



AALBORG UNIVERSITET

Datahåndtering i den kommende nationale biobank

Sundhedsteknologi
10. Semester
2010

Kate Zeberg Olesen
Ditte Haargaard Kopp
10gr1086b

Titel:

Datahåndtering i den kommende nationale biobank

Tema:

Anvendt sundhedsteknologi og informatik

Projektperiode:

P10 Forår 2010

Projektgruppe:

10gr1086b

Forfattere:

Kate Zeberg Olesen
Ditte Haargaard Kopp

Vejleder:

Mette Dencker Johansen

Udgivelser: 5

Sideantal: 133

Vedhæftelser: 1 CD

Afsluttet: 4. Juni 2010

Synopsis:

På baggrund af modsatrettede holdninger mellem SSI og SST angående datahåndteringen i den kommende nationale biobank vil identificeres i dette projekt et kompromis for datahåndteringen og en systemarkitektur skitseres, der benytter sig af standardmetoder. Med inspiration fra UP og ER-modellering udvikles et konceptuelt system til datahåndteringen ved at analysere systemets indhold og metoder, designe systemet ved identificering af specifikke standardmetoder samt evaluere systemkonceptet. Ved et prædefineret tidsinterval håndterer en Javaapplikation kopiering af relevante sundhedsdata fra SST's registre, tildeling af et ID til hvert CPR-nummer og lagring af relevante sundhedsdata i SSI's Biobankregister således dataanonymisering kan foretages. Evaluering af systemet illustrerede at det er muligt at identificere et kompromis samt skitsere en systemarkitektur, der benytter sig af standardmetoder, som kan imødekomme de modsatrettede holdninger til datahåndteringen i den nationale biobank.

Title:

Data management in the upcoming national biobank

Theme:

Applied biomedical engineering and informatics

Project period:

P10 Spring 2010

Project group:

10gr1086b

Authors:

Kate Zeberg Olesen
Ditte Haargaard Kopp

Supervisor:

Mette Dencker Johansen

Publications: 5

Number of pages: 133

Attachments: 1 CD

Finished: June 4th 2010

Synopsis:

A conflicting view between SSI and SST concerning data management in the future Danish national biobank constitutes the background for this project which addresses the identification of a compromise for the data management, and an outlining of the system architecture that utilize standard methods.

With inspiration from UP and ER-modelling a conceptual system is developed for the data-management by means of analysing the system content and methods, designing the system by identifying specific standard methods and evaluating the system concept. At a predefine time interval a Java application handles copying of the relevant data from the registers at SST, assignment of an ID to every CPR-number and storing of the relevant data in the Biobankregister at SSI. Hereby data anonymization is also performed.

Evaluation of the system illustrated, that it is possible to identify a compromise and outline a system architecture, that utilize standard methods, which comply with the conflicting view regarding data management in the danish national biobank.

Forord

Denne rapport er udarbejdet af Kate Zeberg Olesen og Ditte Haargaard Kopp på 10. semester i uddannelsen til civilingeniør i Sundhedsteknologi med speciale i Medicinsk informatik på Aalborg Universitet i foråret 2010. Temaet for dette semester er "Anvendt sundhedsteknologi og informatik".

Rapporten er udarbejdet på baggrund af et projekt omhandlende etablering af en national biobank i Danmark hos Statens Serum Institut (SSI). Projektet er bevilget 121 millioner, heraf 36 millioner fra Videnskabsministeriet og 85 millioner fra Novo Nordisk Fonden. Formålet med den nationale biobank er at styrke forskningen ved at give forskere mulighed for let at koble biologiske prøver med sundhedsdata registreret om patienterne i Sundhedsstyrelsens (SST's) forskellige registre. Derved søges det at forbedre behandlingerne, diagnosticeringen og forebyggelsen af forskellige sygdomme. To personer er ansat til at varetage projektet, og der er på nuværende tidspunkt udelukkende nedfældet overordnede visioner for den nationale biobank. Med udgangspunkt i dette indeholder denne rapport en analyse af mulighederne for etablering af den nationale biobank samt udarbejdelsen af et systemkoncept for datahåndteringen i den nationale biobank. Rapporten henvender sig primært til SSI og SST samt andre med interesse for det nationale biobankprojekt, og den er tænkt til at kunne være et sagligt input i debatten om datahåndtering i projektet med den nationale biobank.

Der rettes en tak til SST, som har givet deres holdning til kende angående projektet med den nationale biobank, samt bidraget med viden om registerforhold. En tak rettes ligeledes til sektordirektør hos SSI, Mads Melbye og biobankchef hos SSI, Kristian Hveem, der har bidraget med viden om visionerne for den nationale biobank, og Mads Melbye har ligeledes hjulpet med evaluering af systemkonceptet. Chefarkitekt hos CSC, John Møller-Jensen takkes ligeledes for hans behjælpelig i forbindelse med evaluering af systemets tekniske indhold. Desuden rettes en tak til Onkologisk afdeling på Vejle Sygehus samt Patologisk Institut på Skejby Sygehus, da disse har leveret viden om specifikke indberetningsdata.

Endvidere takkes SSI for viden om Familie-relations databasen. Endelig takkes vores vejleder, Adjunkt P.hD Mette Dencker Johansen, Aalborg Universitet for kyndig vejledning igennem projektperioden.

Med undtagelse af Mads Melbye, Kristian Hveem og John Møller-Jensen ønskede de interviewede personer udelukkende at tilkendegive deres titel i denne rapport, hvorfor ingen navne er angivet. Vedlagt er en CD, der indeholder materiale til litteraturstudie samt oversigter over dokumentation af de forskellige registre.

Kate Zeberg Olesen

Ditte Haargaard Kopp

Indholdsfortegnelse

1	Forord	7
2	Indledning	15
I	FORANALYSE	17
3	Metodologi til undersøgelse af muligheder for en national biobank i DK	19
3.1	Identificering af operationelle spørgsmål	19
3.2	Metode	21
3.2.1	Visioner for den nationale biobank	21
3.2.2	Status for den nationale biobank	22
3.2.3	Andre instansers kendskab til den nationale biobank	22
3.2.4	Identificering af relevante data	23
3.2.5	Tilgængelighed af relevante data	24
3.2.6	Metodeoversigt	25
4	Muligheder for en national biobank i DK	26
4.1	Visioner for den nationale biobank	26
4.2	Status for den nationale biobank	28
4.3	Andre instansers kendskab til den nationale biobank	28
4.4	Identificering af relevante data	29
4.5	Tilgængelighed af relevante data	32
4.6	Konklusion	33
4.7	Problemformulering	35

5	Løsningsstrategi	36
6	Vision for datahåndteringssystemet	39
7	Metode til systemudvikling	41
7.1	Valg af systemudviklingsmetode	41
7.2	Anvendelse af Unified Process	42
II	KRAV	45
8	Krav	47
8.1	Systemindhold	47
8.2	Use Cases	48
8.3	Aktivitetsdiagrammer med swim lanes	48
8.4	Kravspecifikation	49
8.5	Deploymentdiagram	50
III	ANALYSE	51
9	Systemets dele	53
9.1	Systemarkitektur	53
10	Dataerhvervelse	56
10.1	Tidsinterval for opdatering	56
10.2	Kopiering af relevante data	57
10.3	Tildeling af ID	58
11	Datalagring	59
11.1	Indsættelse af relevante data i Biobankregisteret	59
11.2	Datastruktur	60
11.3	Anonymisering af relevante data	62
12	Datapresentation	63
12.1	Datasikkerhed	63
12.2	Adgang til relevante data	66
13	Aktivitetsdiagram med swimlanes	67

13.1 UC: "Kopier og gem data"	67
13.2 UC: "Visualiser data"	68
IV DESIGN	69
14 Systemets dele	71
14.1 Systemarkitektur	71
15 Dataerhvervelse	73
15.1 Tidsinterval for opdatering	73
15.2 Kopiering af relevante data	73
15.3 Tildeling af ID	74
16 Datalagring	75
16.1 Indsættelse af relevante data i Biobankregisteret	75
16.2 Datastruktur	75
16.3 Anonymisering af relevante data	78
17 Datapræsentation	79
17.1 Datasikkerhed	79
17.2 Adgang til relevante data	80
18 Aktivitetsdiagram med swimlanes	83
18.1 UC: "Kopier og gem data"	83
18.2 UC: "Visualiser data"	84
V EVALUERING og KONSEKVENSVURDERING	85
19 Evaluering	87
19.1 Metode til koncepttest	87
19.2 Resultat	89
19.3 Forbehold	90
19.4 Konklusion	90
20 Konsekvensvurdering	92
20.1 Arbejdsdeling for datahåndtering	92

20.2 Økonomi	94
20.3 Konklusion	94
VI SYNTESE	95
21 Diskussion	97
21.1 Evaluering	97
21.1.1 Resultater	97
21.1.2 Metode	98
21.2 Konsekvensvurdering	98
21.3 Systemudviklingsmetode	98
21.4 Kompromis	99
21.5 Adgang til relevante sundhedsdata	99
22 Konklusion	100
23 Perspektivering	102
LITTERATURLISTE	103
APPENDIKS	107
A Interviewguide	109
A.1 Semi-strukturerede interviews	109
A.1.1 Mads Melbye	109
A.1.2 Kristian Hveem	110
A.1.3 Sundhedsstyrelsen	110
A.2 Informantinterviews	111
A.2.1 Onkologisk afdeling, Vejle Sygehus	111
A.2.2 Patologisk institut	111
A.2.3 Landspatientregisteret	111
A.2.4 Fødselsregisteret	112
A.2.5 Familie-relations databasen	112
A.2.6 Forskerservice	112
A.2.7 Sundhedsstyrelsen	112

B	Semi-strukturerede interviews	113
B.1	Mads Melbye	113
B.2	Kristian Hveem	114
B.3	Sundhedsstyrelsen	115
C	Informantinterviews	118
C.1	Onkologisk afdeling, Vejle Sygehus	118
C.2	Patologisk institut	119
C.3	Landspatientregisteret	119
C.4	Fødselsregisteret	119
C.5	Familie-relations databasen	119
C.6	Forskerservice	120
C.7	Sundhedsstyrelsen	120
D	Unified Process	121
E	Dataformat	125
E.1	SKS og SNOMED	125
F	DBMS	126
G	Korrespondence ang. Evaluering	128
G.1	Mads Melbye	128
G.2	Kristian Hveem	129
G.3	Ansvarlig for SST's dokumentationsenhed	130
G.4	John Møller-Jensen	132

Indledning

I Danmark bliver mange sygdomme diagnosticeret ved hjælp af biologiske prøver (blodprøver og vævsprøver), der gemmes i forskellige biobanker placeret i hele landet. (Novo Nordisk Fonden, 2009) Prøverne bliver taget ved fødsel (f.eks. hælprøver) og ved indlæggelser, hvor der kræves prøveudtag til diagnosticering, og indeholder derfor informationer om fødselstilstand og sundhedstilstand (rask/syg), som kan bruges i forskningen om forebyggelse og behandling af sygdomme (Landegren, 2002; Jonsson og Landegren, 2004). Biobankerne indeholder ligeledes metadata om prøverne, der beskriver, hvad der kan findes i prøverne. Yderligere indberettes forskellige sundhedsdata om patienten, via patientens CPR-nummer, fra sygehusenes Patientadministrative Systemer (PAS) til bl.a. Landspatientregisteret, heriblandt patientens diagnose, patologisvar samt den valgte behandling af sygdommen (Ministeriet for Sundhed og Forebyggelse, Sundhedsstyrelsen, Danske Regioner og Finansministeriet, 2009; Statens Institut for folkesundhed, 2007). Ligeledes indberettes det, hvis patienten dør eller en ny diagnose findes. Disse sundhedsdata er ligeledes tilgængelige i sygehusenes Elektroniske Patientjournaler (EPJ).

Andre europæiske lande har etableret projekter, hvor biologiske prøver og sundhedsdata samles i en fælles national biobank. I Norge blev der i 1984 taget initiativ til en national biobank, der inkluderer sundhedsdata fra 120.000 personer, hvoraf 80.000 har afgivet blodprøver. Formålet med biobanken er at give forskerne mulighed for at få yderligere indsigt i sygdomme og sundhed ved at kombinere sundhedsdata og biologiske prøver med henblik på at undersøge kombinationen af genetisk variation, livsstil og miljøpåvirkning. (HUNT Biosciences AS, 2008) Sverige er et andet eksempel på et skandinavisk land, der har etableret en national biobank. Biobanken inkluderer 50-100 millioner biologiske prøver, og hvert år bliver yderligere 3-4 millioner biologiske prøver lagret. Formålet med biobanken er at forbedre diagnosticeringen af sygdomme, kontrollere epidemiske og/eller infektiøse sygdomme, forbedre undervisning af medicinsk personale og ikke mindst den medicinske forskning. (SWEGENE, 2003) Også England tog i 2006 initiativ til at lave en national biobank, der på langt sigt skal indeholde 500.000 biologiske prøver fra folk i aldersgrup-

pen 40-69 år. Formålet med biobanken er at forbedre forebyggelse, diagnose og behandling af forskellige livstruende sygdomme, herunder bl.a. kræftsygdomme, diabetes og demens, ved at kombinere sundhedsdata med biologiske prøver. (UK Biobank Coordinating Centre, 2006)

Danmark har ikke en fælles biobank, hvorfor de biologiske prøver er placeret i forskellige biobanker landet over. Ligeledes er de registrerede sundhedsdata indberettet til forskellige registre. Dermed udnyttes det potentiale, som prøverne og de mange sundhedsdata besidder, ikke fuldt ud, idet forskerne ikke har let adgang til netop de data, der kan bidrage til en forbedring af forskningen i sygdomme og behandlinger. Samtidig forhindres overblikket over mængden og sammensætningen af prøver.

Følges principperne fra andre europæiske landes nationale biobanker kan det danske problem afhjælpes ved at samle alle blod- og vævsprøver samt sundhedsdata i en national biobank. Derved optimeres forskernes mulighed for at uddrage information fra såvel biologiske prøver som sundhedsdata. Etablering af en national biobank giver forskerne mulighed for at lave en kobling mellem blod- og vævsprøver med patologisvar, diagnose, behandling og behandlingseffekt for hver enkelt patient, hvorved der opstår mulighed for at identificere og bl.a. genetisk karakterisere patientgrupper samt deres positive eller negative respons på en given behandling eller forebyggelse. Dermed kan forskernes adgang til data og prøver på lang sigt gavne klinikerne og i sidste ende patienterne, idet behandlingsmetoderne forbedres og effektiviseres, og bedre resultater af behandlinger opnåes. (Novo Nordisk Fonden, 2009)

Initiativet til en national biobank i Danmark er taget af Statens Serum Institut (SSI), der råder over en større mængde biologiske prøver og som derfor skal være ansvarlig for projektet. Dog er det udelukkende formål og visioner der er nedfældet, se afsnit 4.1 på side 26, og kun to personer er på nuværende tidspunkt ansat på projektet, hvorfor det må antydes at projektet er i sin startfase. Da etableringen af en national biobank i Danmark må anses for at være en udfordring pga. mængden af biologiske prøver og sundhedsdata placeret i forskellige biobanker og registre, samt det faktum at projektet kun er i den begyndende fase, ønskes det at undersøge, hvilke muligheder der er for etablering af en national biobank i Danmark. Dette leder til følgende initierende problem:

Hvad er mulighederne for en national biobank i Danmark, hvorfra det er muligt at koble biologiske prøver med relevante sundhedsdata?

Del I

FORANALYSE

I foranalysen analyseres mulighederne for en national biobank i Danmark ved hjælp af metoderne litteraturstudie, informantinterviews, semi-strukturerede interviews og brainstorming.

Metodologi til undersøgelse af muligheder for en national biobank i DK

Det initierende problem blev søgt besvaret ved at nedbryde det til operationelle spørgsmål, som blev besvaret i den videre analyse. Forskellige metoder blev anvendt til besvarelse af de operationelle spørgsmål.

3.1 Identificering af operationelle spørgsmål

Det initierende problem kan opdeles i to, idet den aktuelle progression og status for planer samt forudsætningers tilstedeværelse tilsammen må antages at udgøre fundamentet for mulighederne for en national biobank i Danmark. De to områder er opsummeret i nedenstående:

- Progressioner og planer for en national biobank
- Forudsætninger for en national biobank

Progressioner og planer

Indsigt i progressioner og planer for en national biobank kunne opnås ved først at få indblik i visionerne for den nationale biobank. Ydermere var det nødvendigt at undersøge status for projektet samt andre instansers kendskab til den nationale biobank, idet det er relevant at vide om datahåndteringsprocessen allerede er igangsat. Herfra udledtes tre operationelle spørgsmål:

- Hvad er visionen for den nationale biobank?
- Hvad er status for den nationale biobank?
- Hvilket kendskab har andre instanser til den nationale biobank?

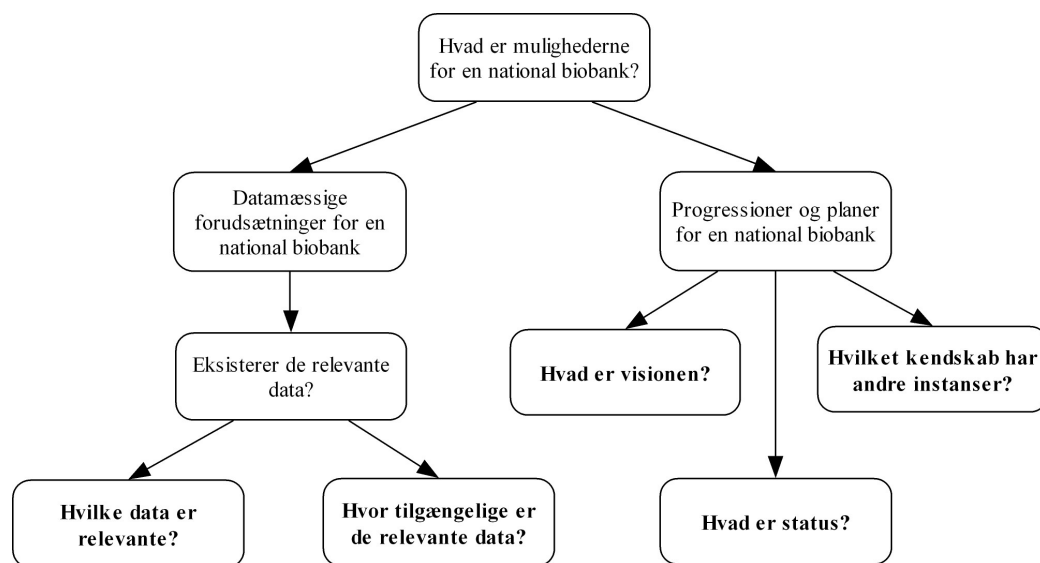
Forudsætninger

En vigtig forudsætning for en national biobank er, at de relevante sundhedsdata og biologiske prøver eksisterer. SSI besidder allerede en stor del af de biologiske prøver i Danmark, hvorfor en stor del af udfordringen i opførelsen af den nationale biobank består i at indhente sundhedsdata. Derfor fokuseres der i dette projekt på datahåndteringsdelen i etableringen af den nationale biobank. Formålet med den nationale biobank er, at den skal indeholde data, der kan kobles med biologiske prøver. Derfor skulle disse data identificeres og efterfølgende skulle det undersøges, hvorvidt disse data er tilgængelige. På baggrund af dette blev to operationelle spørgsmål undersøgt:

- Hvilke data er relevante?
- Hvor tilgængelige er de relevante data?

Operationelle spørgsmål

Figur 3.1 illustrerer udledningen af de tre operationelle spørgsmål.



Figur 3.1: Mulighederne for en national biobank i Danmark identificeredes vha. udledningen af fem operationelle spørgsmål. De operationelle spørgsmål er skrevet med **fed** skrift.

Nedenstående lister de fem udledte operationelle spørgsmål:

- Hvad er visionen for den nationale biobank?
- Hvad er status for den nationale biobank?
- Hvilket kendskab har andre instanser til den nationale biobank?
- Hvilke data er relevante?

- Hvor tilgængelige er de relevante data?

En besvarelse af de operationelle spørgsmål kan samlet hjælpe til at udlede mulighederne for en national biobank i Danmark.

3.2 Metode

De fem operationelle spørgsmål blev besvaret ved anvendelse af forskellige metoder.

3.2.1 Visioner for den nationale biobank

Kendskab til visionerne for den nationale biobank i Danmark blev opnået ved at lave et litteraturstudie samt et semi-struktureret interview med de projektansvarlige fra SSI.

Litteraturstudie

For at få indsigt i visionerne for den nationale biobank blev officielle danske hjemmesider med relation til det nationale biobankprojekt undersøgt for publikationer og information om den nationale biobank. Da initiativet til den nationale biobank foreligger hos SSI blev deres hjemmeside undersøgt for publikationer ang. den nationale biobank. Novo Nordisk Fonden blev udvalgt, da denne har bidraget til størstedelen af finansieringen og derfor må formodes at have en interesse i projektet, hvilket ligeledes gør sig gældende for Videnskabsministeriet, da de bidrog med den resterende del af finansieringen. De valgte søgeord var "biobank i Danmark" og "national biobank". Links til publikationer, der fremkom ud fra søgeordene, blev studeret for relevans og udvalgt ud fra konsensus i projektgruppen.

Semi-strukturerede interviews

Der blev foretaget kvalitative interviews med henblik på at få information om visionerne for den nationale biobank. Endvidere blev det valgt at have et semi-struktureret design, da spørgsmålene der ønskedes besvaret ikke skulle besvares i den specificerede rækkefølge, og ligeledes skulle der være mulighed for yderligere og opklarende spørgsmål undervejs i interviewene. Som beskrevet i indledningen er initiativet til en national biobank fra SSI, hvorfor telefoniske interviews blev foretaget med Mads Melbye (sektordirektør, SSI) og Kristian Hveem (Biobankchef, SSI), da de som projektansvarlige ansås for at være relevante ift. at få klarlagt visionen for en national biobank i Danmark. Interviews blev optaget og efterfølgende transskriberet. Ud fra besvarelserne blev relevant indhold, der kunne bruges til besvarelse af det operationelle spørgsmål, uddraget. Interviewspørgsmålene fokuserede på visionerne og strategierne for den nationale biobank og kan ses i interviewguiden i Appendiks A.

3.2.2 Status for den nationale biobank

For at få indsigt i status for den nationale biobank i Danmark blev der foretaget et semi-struktureret interview med en projektansvarlig fra SSI.

Semi-struktureret interview

Der blev foretaget kvalitative interviews med henblik på at få information om status for den nationale biobank. Ligesom i det ovenstående blev det valgt at have et semi-struktureret design. Interviewet blev foretaget med Mads Melbye, da han er initiativtager til projektet, ved telefonisk henvendelse. Interviews blev optaget og efterfølgende transskriberet. Ud fra besvarelserne blev relevant indhold, der kunne bruges til at beskrive status for den nationale biobank, uddraget. De stillede spørgsmål omhandlede status for den nationale biobank og specifikt status for datahåndteringsprocessen og kan ses i interviewguiden i Appendiks A.

3.2.3 Andre instansers kendskab til den nationale biobank

Kendskab til den nationale biobank i Danmark blev afdækket ved at lave et litteraturstudie samt et semi-struktureret interview med en ansvarlig for de forskellige eksisterende sundhedsdata i Danmark.

Litteraturstudie

For at vurdere, hvorvidt andre instanser end SSI havde kendskab til og indsigt i planer og progressioner for den nationale biobank i Danmark, blev officielle hjemmesider for Sundhedsstyrelsen (SST), Novo Nordisk Fonden og Videnskabsministeriet undersøgt for information om det nationale biobankprojekt. SST's hjemmeside blev undersøgt for publikationer angående den nationale biobank, da de på nuværende tidspunkt råder over størstedelen af alle sundhedsdata, og derfor er en central kilde til data til den nationale biobank. Novo Nordisk Fonden og Videnskabsministeriet blev udvalgt på samme grundlag som beskrevet i afsnittet om visioner for den nationale biobank, se afsnit 4.1. De valgte søgeord var "biobank i Danmark" og "national biobank". Links til publikationer, der fremkom ud fra søgeordene, blev studeret for relevans og udvalgt ud fra konsensus i gruppen.

Semi-struktureret interview

Endnu et kvalitativt interview blev foretaget med henblik på at få information om andre instansers kendskab til den nationale biobank. Det blev ligeledes valgt at have et semi-struktureret design. Interviewet blev foretaget telefonisk med en ansvarlig for de eksisterende registre fra SST, da de er en central kilde til data, hvorfor dennes viden om det nationale biobankprojekt eller mangel herpå var en indikator for kendskabet til den nationale

biobank. Interviewet blev optaget og efterfølgende transskriberet. Ud fra besvarelsene blev relevant indhold, der kunne bruges til at beskrive andre instansers kendskab til den nationale biobank, uddraget. Spørgsmålene omhandlede SST's kendskab til den nationale biobank og deres holdning til projektets formål. Yderligere blev der stillet spørgsmål om SST's rolle i forbindelse med den nationale biobank, se Appendiks A.

3.2.4 Identificering af relevante data

Identificering af de relevante data blev opnået ved at foretage et litteraturstudie til at indhente viden om visionerne for data indeholdt i den nationale biobank samt data indeholdt i andre landes nationale biobanker. Efterfølgende blev der, med udgangspunkt i litteraturstudiet, lavet brainstorming til udledning af, hvilke data der var relevante i en dansk national biobank.

Litteraturstudie

For at få kendskab til visionerne for dataindhold i den nationale biobank blev officielle hjemmesider fra SSI, SST, Novo Nordisk Fonden og Videnskabsministeriet undersøgt for publikationer angående dataindhold i den nationale biobank, idet SSI er initiativtagerne på projektet, SST er dataleverandør og Novo Nordisk Fonden og Videnskabsministeriet er finansierende instanser. De valgte søgeord var "biobank i Danmark" og "national biobank". Idet Danmark ikke har haft nogle nationale biobankprojekter, var viden om data indeholdt i andre landes biobanker interessante. Kriterierne for udvælgelsen af andre landes nationale biobanker var, at det overordnede koncept for de udenlandske nationale biobanker var sammenlignelig med konceptet for den danske biobank som beskrevet i introduktionen. Derfor blev officielle hjemmesider og publikationer fra biobanker i Norge, Sverige og England studeret, hvori biobankernes formål samt dataindhold var beskrevet, for at få inspiration om, hvilke data der er relevante i den nationale biobank i Danmark. Publikationerne blev valgt ud fra indholdsfortegnelsens relevans i forhold til dataindhold og formål for biobanken, og efterfølgende blev de forskellige informationer diskuteret samt sorteret, og kerneinformation blev uddraget ud fra konsensus i gruppen.

Brainstorming

Visionerne for dataindholdet i den nationale biobank blev undersøgt, da det vurderedes at disse kunne specificere, hvilke sundhedsdata den nationale biobank bygger på. Til trods for at formålet med de forskellige biobanker ikke er ens, blev det antaget, at alle nationale biobanker har et fælles etableringsgrundlag (biologiske prøver og dertilhørende sundhedsdata) og dermed også nogle fælles data, der er nødvendige i enhver national biobank. Derudover blev eventuelle forskelle i formålene med biobankerne brugt som inspiration til

indholdet i den nationale biobank i DK. Afdækning af relevante sundhedsdata i en dansk national biobank blev derfor foretaget vha. brainstorming på baggrund af viden om data indeholdt i andre landes nationale biobanker samt viden om medicinsk informatik opnået gennem uddannelsen. Brainstorming blev valgt som metode for at sikre, at alle aspekter angående relevante sundhedsdata var inkluderet, og at der var taget forbehold for specifikke danske forhold. Brainstorming blev udført løbende af gruppens medlemmer, og meta-viden opnået gennem brainstorming blev filtreret ud fra formålet med den nationale biobank i Danmark.

3.2.5 Tilgængelighed af relevante data

Indhentning af viden om tilgængeligheden af relevante data kunne opnås ved at udføre informantinterviews med forskellige sygehusafdelinger og registre.

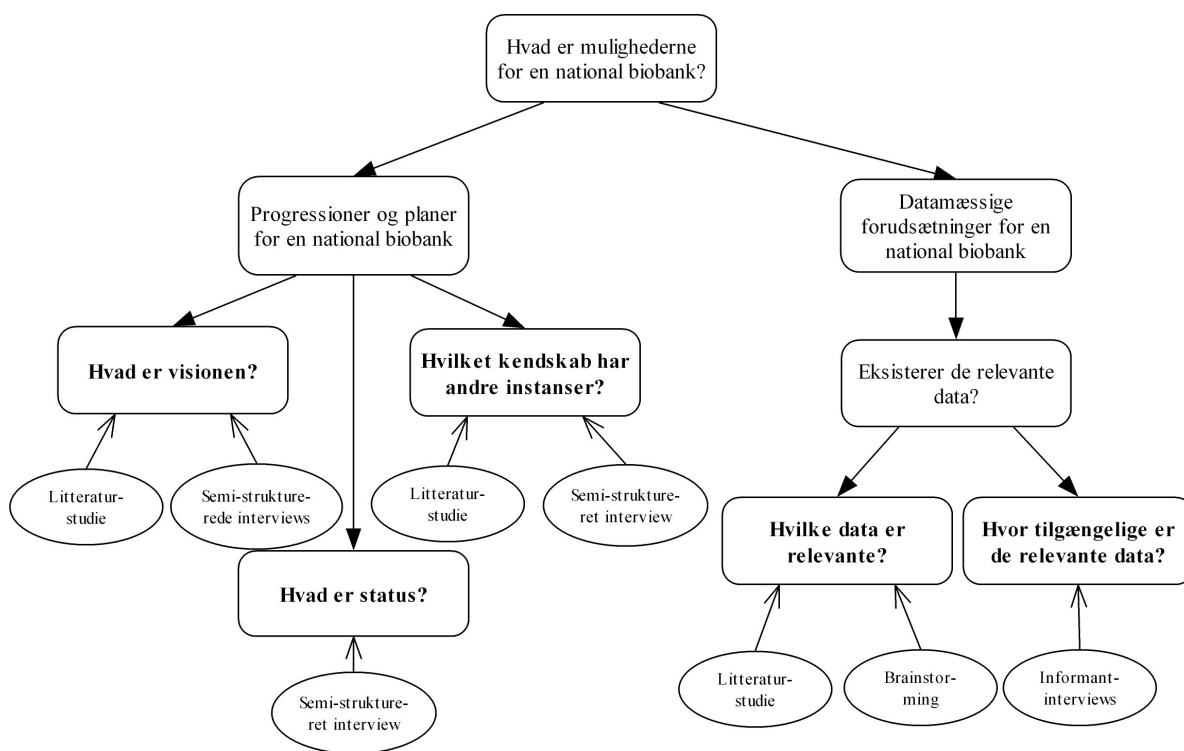
Informantinterviews

Da der ønskedes specifik viden om tilgængeligheden af relevante data, blev der udarbejdet nogle få konkrete spørgsmål, hvorfor den metodiske fremgangsmåde var strukturerede informantinterviews af kvalitativ karakter. Viden om specifikke danske registreringsforhold og dermed tilgængelighed af relevante sundhedsdata blev opnået ved kontakt med nøglepersoner, med konkret viden om disse forhold, fra både leverandører (sygehusafdelinger) og modtagere (SST's registre) af data. Sygehusafdelingerne blev udvalgt, idet de registrerer alt data om patienterne. Der var to kriterier for valg af sygehusafdelinger som begge skulle være opfyldt: korte og lange patientforløb samt et større forbrug af både blod- og vævsprøver, og viden om data om biologiske prøver. Derfor blev Onkologisk afdeling i Skejby kontaktet, idet en onkologisk afdeling opfylder kriterierne om både korte og lange patientforløb og ligeledes har et stort forbrug af blod- og vævsprøver. For at opnå viden om, hvorvidt data om blod- og vævsprøver gemmes, blev Klinisk Biokemisk afdeling i Aalborg og Patologisk Institut i Skejby ligeledes kontaktet. Nøglepersonerne fra de valgte afdelinger blev udvalgt ud fra en forventning om, at de har viden om specifikke registreringsforhold.

Registrene blev kontaktet for at undersøge, om de relevante sundhedsdata ligeledes bliver indberettet til forskellige registre. Nøglepersoner fra registrene blev valgt ud fra en vurdering af, hvem der kunne udtale sig kvalificeret om datatilgængelighed, hvorfor nøglepersoner fra Landspatientregisteret blev kontaktet. Spørgsmålene omhandlede, hvorvidt diagnose, behandling og medicinske undersøgelser registreres i Landspatientregisteret. Ydermere kontaktes nøglepersoner fra Fødselsregisteret og Familie-Relations Databasen kontaktet, der blev spurgt om, hvorfra deres data udtrækkes. Dataindsamlingen foregik via telefonisk henvendelse, hvor hvert interview blev optaget og efterfølgende transskriberet. Ud fra besvarelserne blev relevant indhold, der kunne bruges til besvarelse af det operationelle spørgsmål, uddraget. De forskellige spørgsmål kan ses i interviewguiden i Appendiks A.

3.2.6 Metodeoversigt

I figur 3.2 ses de identificerede metoder til at undersøge hver af de fem operationelle spørgsmål.



Figur 3.2: De operationelle spørgsmål med dertilhørende metoder, der skal anvendes til besvarelse af de fem operationelle spørgsmål, er illustreret som ovaler. De forskellige pile illustrerer henholdsvis metoder og operationelle spørgsmål.

Kapitel 4

Muligheder for en national biobank i DK

4.1 Visioner for den nationale biobank

Litteraturstudie

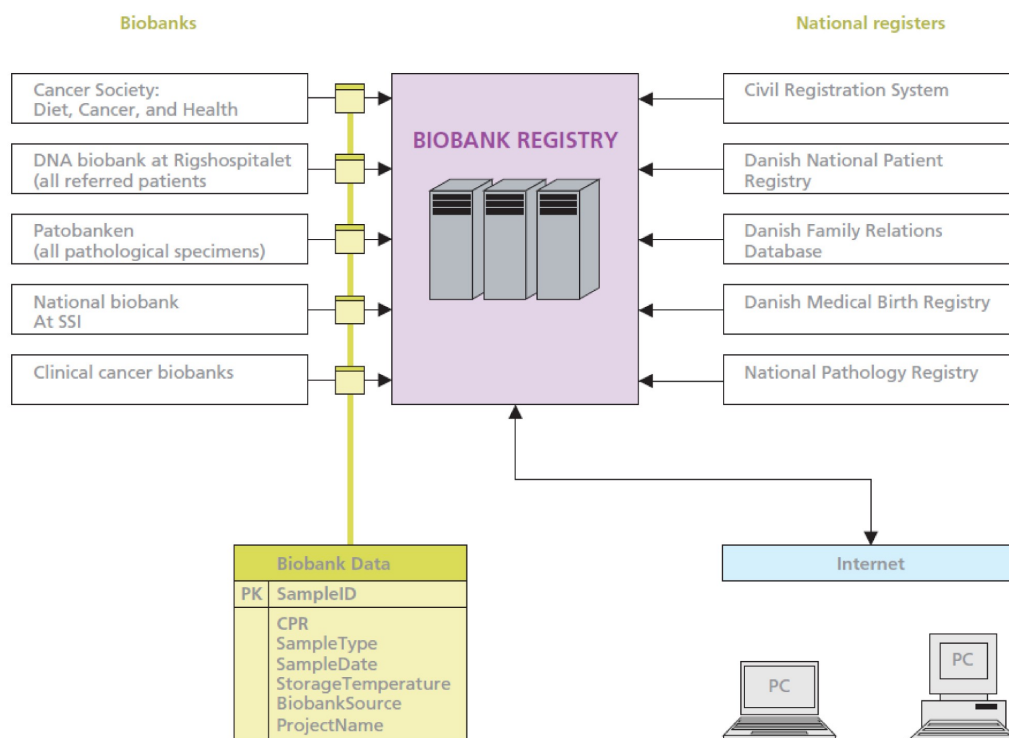
De officielle hjemmesider fra SSI og Videnskabsministeriet har udelukkende links til publikationer fra Videnskabsministeriet, der overordnet beskriver formålet med den nationale biobank, mens Novo Nordisk Fonden har publiceret materiale, hvor visioner for den nationale biobank beskrives af Mads Melbye.

Ved undersøgelse af visionerne for den nationale biobank ses det, at visionerne inddeles i tre dele: et Biobankregister, en Prøvebank og et Koordinerende center (Novo Nordisk Fonden, 2009; Videnskabsministeriet, 2009). Visionerne blev nedfældet, da projektet i 2009 ansøgte om økonomisk støtte hos Novo Nordisk Fonden.

Biobankregister

Biobankregisteret skal indeholde stamdata, der indhentes ved regelmæssig kobling til allerede eksisterende biobanker såsom Patobanken, De Kliniske Kræftbiobanker og SSI's biologiske prøver. Derudover indhentes data fra nationale registre såsom Landspatientregisteret og Patologiregisteret vha. CPR-nummer. Tilsammen giver disse data mulighed for at uddrage, hvordan det er gået de forskellige patienter. Ét af formålene med Biobankregisteret er, at give forskere let adgang til data ved at lave en webportal, hvor aggregerede data, der ikke er personhenførbare, kan tilgås. Hermed kan forskere i Danmark og internationalt få online indblik i dele af Biobankregisteret og danne sig et overblik over, hvad der potentielt set findes af materiale på patienter, der har udviklet forskellige sygdomme. Samtidig har en betænkning på biobankområdet tilgodeset individets ret til medbestemmelse over egne prøver, og der er derfor oprettet et register for personer, der ikke ønsker at deres materialer bliver brugt til forskning (Dillner, 2002). Figur 4.1 på næste side illustrerer en

skematisk fremstilling af Biobankregisteret som Mads Melbye forestiller sig det færdige resultat. (Novo Nordisk Fonden, 2009)



Figur 4.1: Skematisk fremstilling af Biobankregisteret som det gerne skulle komme til at fungere i henhold til SSI. (Novo Nordisk Fonden, 2009)

Prøvebank

Prøvebanken etableres på SSI og en af opgaverne bliver at registrere og opbygge brugervenlige arkiveringssystemer til de eksisterende biologiske prøver, der idag opbevares i store frysevarerhuse. En anden opgave bliver at modtage nyankomne prøver til instituttet, der vurderes til at være ca. 200.000 om året. (Novo Nordisk Fonden, 2009)

Koordinerende center

Det koordinerende center skal håndtere opbygningen af Biobankregisteret samt Prøvebanken og herefter stå for varetagelse af funktionerne i disse på daglig basis. Yderligere skal forskere kunne søge rådgivning i spørgsmål, der relaterer sig til brug af Biobankregisteret og Prøvebanken. (Novo Nordisk Fonden, 2009)

Semi-strukturerede interviews

Ud fra interview med Kristian Hveem og Mads Melbye, se Appendiks B.1 og B.2, blev det klargjort, at visionerne og den fysiske beliggenhed af den nationale biobank er på plads, og det vides, hvorfra data skal indsamles.

Delkonklusion

Den nationale biobank kommer til at bestå af tre dele: Biobankregisteret, Prøvebanken og det Koordinerende center. Biobankregisteret skal indeholde sundhedsdata, Prøvebanken skal opbevare biologiske prøver, mens det Koordinerende center skal varetage funktionerne i Biobankregisteret og Prøvebanken. Ydermere skal der oprettes en online webportal, hvorfra forskere kan få adgang til aggregerede sundhedsdata.

4.2 Status for den nationale biobank

Semi-struktureret interview

De semi-strukturerede interviews med Kristian Hveem og Mads Melbye har bidraget med indsigt i status for projektet. På nuværende tidspunkt er Kristian Hveem og Mads Melbye de eneste personer ansat på projektet, idet det idelukkende er det overordnede systemkoncept, der er i fokus. Initielt arbejdes der med at planlægge konstruktionen/opbygningen af den fysiske biobank, som er en lang og tidskrævende proces, hvor bl.a. landets øvrige biobanker skal inkluderes og rådføres med. Derfor er det besluttet at arbejdet med Biobankregisteret og selve datahåndteringen kommer i anden fase. Dog er det hensigten, at data skal indhentes fra SST's registre, og diskussioner mellem SST og SSI angående dette er allerede indledt, se interview med Mads Melbye i Appendiks B.1. På baggrund af dette kan det udledes, at der på nuværende tidspunkt ikke er nogen konkrete planer for, hvordan erhvervelse samt håndtering af den store datamængde kan og i givet fald skal designes og implementeres. Se interviews med Mads Melbye og Kristian Hveem i Appendiks B.1 og B.2.

Delkonklusion

Den nationale biobank er fastlagt med hensyn til arkitektur. Der arbejdes dog indledningsvist kun med opbygningen af Prøvebanken, mens opbygningen af Biobankregisteret udskydes til Prøvebanken er på plads. Dog ønskes det at indhente data fra SST's registre.

4.3 Andre instansers kendskab til den nationale biobank

Litteraturstudie

Ved at undersøge SST's, Videnskabsministeriets og Novo Nordisk Fondens officielle hjemmesider blev det klart, at SST ikke har nogen publikationer eller links angående den nationale biobank. Derimod har Videnskabsministeriet og Novo Nordisk Fonden begge publikationer omhandlende den nationale biobank dateret tilbage til det økonomiske bevillingstid-

spunkt, hvilket derfor hentyder til, at kun de økonomisk involverede parter har kendskab til projektet med den nationale biobank.

Semi-struktureret interview

Det semi-strukturerede interview med SST synliggjorde, at den registeransvarlige hos SST kender til projektet med den nationale biobank, og at repræsentanter fra SST deltager i møderne omkring de overordnede visioner. Som beskrevet ovenfor i visionerne og ifølge telefoninterview med Mads Melbye er et vigtigt perspektiv for Biobankregisteret at indhente sundhedsdata fra SST's registre, men den strategi er SST ikke enige i. SST understreger, at de har spidskompetencer indenfor registerområdet, og at de derfor besidder viden om dataindhold og datadefinitioner, idet de også varetager kommunikationen med sygehusene angående dataindberetning. Disse kompetencer bevirker, at SST har ekspertviden om den overvejende del af alle registre i Danmark, og de mener derfor ikke, at det er hensigtsmæssigt at andre instanser skal overtage deres kernefunktioner. SST understreger, at deres fokus er at producere data af høj kvalitet og have indgående viden om disse data, hvorimod fokus hos projektet med den nationale biobank omhandler at lette forskeres adgang til data. Dog kan SST se det fornuftige i, at SSI får adgang til data omhandlende patologiske prøver, så de får mulighed for at koble deres biologiske prøver med relaterede sundhedsdata. Ud fra dette kan det udledes, at SST ikke finder det hensigtsmæssigt, at lade andre råde over deres data, da de besidder helt centrale kompetencer inden for registerområdet. I Appendiks B.3 kan interview med SST læses.

Delkonklusion

Videnskabsministeriet, Novo Nordisk Fonden og SST har alle kendskab til projektet. SST har udtrykt skepsis omkring dataindholdet i Biobankregisteret, idet de har ekspertviden om deres registre samt indholdet i disse, og dermed den største kompetence i fht. vejledning af forskere. Der er derfor modsatrettede holdninger mellem SSI og SST angående indhentning af data, som kan forhindre, at Biobankregisteret kommer til at fungere efter visionerne.

4.4 Identificering af relevante data

Litteraturstudie

Ved at studere forskellige officielle hjemmesider for viden om dataindholdet i den nationale biobank samt publikationer fra nationale biobanker i Norge, Sverige og England er det blevet klarlagt, at formålet med biobanken har betydning for valg af data indeholdt i biobankerne.

Visionerne for den nationale biobank skrevet af Mads Melbye er den eneste publikation,

der udtaler sig om dataindholdet. Heraf fremgår det, at sundhedsdata vedrørende CPR-nummer, diagnose, familieforhold og behandling skal indgå i Biobankregisteret. Derudover er det illustreret, hvilke registre der ønskes adgang til for at imødegå visionerne for den nationale biobank.

Den nationale biobank i Norge indeholder mange typer af data, der er indhentet fra forskellige registre, såsom resultater af medicinske undersøgelser, laboratoriedata, data om sundhedsstatus, familiehistorik, levestandard, livsstil og medicinering. Disse data er nødvendige i den nationale biobank i Norge, idet formålet med biobanken er at undersøge kombinationen af genetisk variation, livsstil og miljøpåvirkning. (HUNT Biosciences AS, 2008)

Den nationale biobank i Sverige indeholder udelukkende data omhandlende de biologiske prøver. Disse data er relevante at registrere, idet de giver overblik over de biologiske prøver, og registreringen kan derfor bidrage med en forbedring og effektivisering af diagnosticeringen samt forskningen i forskellige sygdomme. (SWEGENE, 2003)

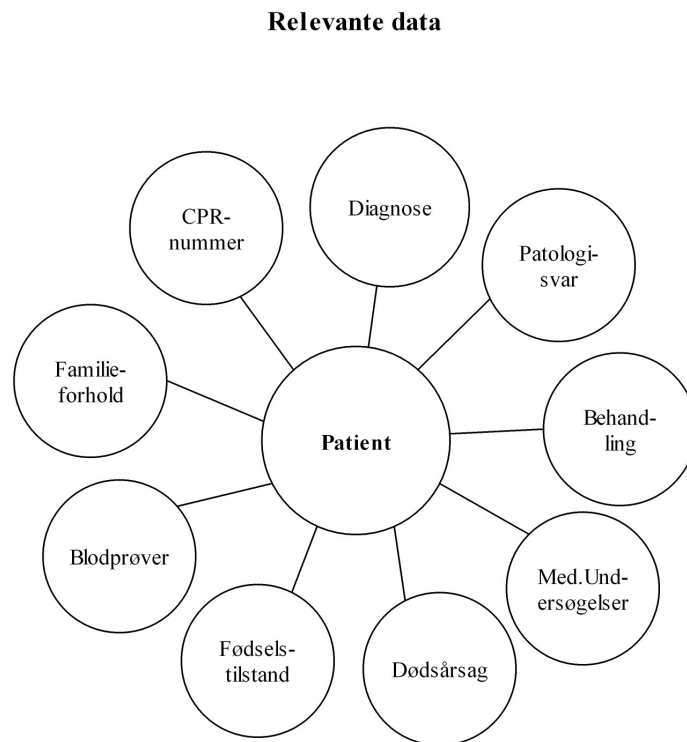
I England er der ligeledes etableret en national biobank til det formål at forbedre forebyggelse, diagnose og behandling af forskellige og ligeledes livstruende sygdomme. Data til biobanken indhentes ved at lave et spørgeskema omhandlende følgende områder: socio-demografi, profession, livsstil, psykologisk tilstand, kognitiv funktion, familiære sygdomme, anamnese og generel sundhedstilstand. (UK Biobank Coordinating Centre, 2006) I Norge og Sverige kan patientdata tilknyttes et personnummer, hvilket ikke er muligt i England.

Brainstorming

Da den nationale biobank i Danmark har til formål at identificere og evt. genetisk klassificere de patientgrupper, der responderer på en specifik behandling eller forebyggelse, samt de patientgrupper der ikke har gavn af behandlingen, bør data omhandlende disse områder være repræsenteret. Med udgangspunkt i formålet og kendskab til visionerne for den nationale biobank i Danmark, samt viden om medicinsk informatik er der lavet brainstorming til identificering af relevante data.

For at give forskerne mulighed for at undersøge hele befolkningen eller udvalgte patientgrupper er CPR-nummeret relevant data, da alle patientoplysninger er bundet hertil. Ligeledes er det nødvendigt at have kendskab til patientens familieforhold, hvis det ønskes at undersøge, hvorvidt en patient er genetisk disponeret for forskellige sygdomme. Endvidere er data vedrørende patientens fødselstilstand fordelagtige, idet de giver forskerne mulighed for at se om patientens sygdom kan relateres tilbage til fødslen. Behandlingsdata er ligeledes relevante parametre, idet forskerne har brug for disse, for at kunne undersøge effekten af forskellige behandlinger. For yderligere at få indsigt i, hvorfor nogle patienter med samme sygdom responderer på en given behandling, mens andre ikke gør, er det en fordel at få kendskab til data om de biologiske prøver eller det patologisvar, der ofte ligger til grund for valget af behandling. I figur 4.2 på modstående side ses de relevante

sundhedsdata der skal være en del af den nationale biobank:



Figur 4.2: Brainstorming over de relevante data, der skal indgå i den nationale biobank.

Ovenstående sundhedsdata identificeret ved hjælp af brainstorming illustrerer de sundhedsdata, projektgruppen finder relevante for den nationale biobank for at visionerne for denne kan opfyldes. Data omhandlende patientens CPR-nummer, familieforhold, behandling og diagnose er dog direkte uddraget fra Mads Melbyes notat omhandlende visionerne for den nationale biobank.

Den nationale biobank fra Norge indeholder ligeledes data om levestandard, livsstil og medicinering. Disse data vil ikke blive inkluderet i den nationale biobank i Danmark, idet denne ikke fokuserer på at undersøge kombinationen af livsstil og miljøpåvirkning. Samtidig er data omhandlende socio-demografi, profession, psykologisk tilstand og kognitiv funktion, der er indeholdt i den engelske biobank, ikke relevante data, da de ikke bidrager til formålet med den nationale biobank i Danmark.

Delkonklusion

Indblik i andre landes nationale biobanker samt visionerne for den nationale biobank giver ideer til hvilke relevante data, der kan inkluderes i den nationale biobank. Derudover kan brainstorming, på visionerne for den nationale biobank, anvendes til at identificere de relevante sundhedsdata. Den nationale biobank skal derfor besidde følgende data, der skal være tilgængelige for forskere:

- CPR-nummer
- Medicinske undersøgelser
- Diagnose
- Behandling
- Dødsårsag
- Patologisvar
- Fødselstilstand
- Familieforhold
- Blod- og vævsprøver

4.5 Tilgængelighed af relevante data

Informantinterviews

Sygehusafdelinger

Alle data om patienten heriblandt patientens CPR-nummer, navn, bopæl, medicinske undersøgelser, diagnose, behandling og dødsårsag registreres med koder i sygehusets PAS, når en patient kommer i kontakt med det danske sygehusvæsen. Hver gang patienten bliver stillet en ny diagnose opdateres denne i PAS sammen med den valgte behandling via patientens CPR-nummer. Er behandlingen længerevarende, opdateres diagnosen løbende, så der er kendskab til patientens sygdomstilstand (forbedret/status quo/forværret). Skal patienten indgå i et kontrolforløb, inden denne kan erklæres helt rask, gives diagnosen en kontrolkode. Hvis patienten erklæres rask eller dør, afsluttes patientforløbet, og ud fra denne viden samt diagnose kan behandlingsoutcome udledes.

Ved undersøgelse af vævsprøver på patologisk afdeling registreres patologisvar i en patologisk database tilknyttet sygehuset. Herfra indberettes de til Patologiregisteret, hvorfra data er tilgængelig for andre registre såsom Cancerregisteret.

Se svar fra informantinterviews i Appendiks C

Registre

Ved kontakt med Landspatientregisteret blev det klart, at alle data der registreres i sygehusenes PAS indberettes til Landspatientregistret via patientens CPR-nummer. Fødsler registreres også i sygehusenes Landspatientregisteret, og indberettes derefter til Fødselsregisteret, medmindre de foretages som hjemmefødsler. Ved hjemmefødsler indsendes en blanket til Fødselsregisteret, som herefter udtrækker data fra CPR-registeret. Dør patienten,

kan dette ses i Dødsårsagsregisteret. Både Landspatientregisteret, Dødsårsagsregisteret, Patologiregisteret, CPR-registeret og Fødselsregisteret hører under SST. Dette betyder, at relevante sundhedsdata, med undtagelse af familieforhold, er tilgængelige fra SST og skal derfor udtrækkes herfra for at indgå i Biobankregisteret. Derimod bliver data omhandlende familieforhold registreret i Familie-relationsdatabasen, som tilhører SSI, og er dermed allerede tilgængelige for Biobankregisteret. Data om blod- og vævsprøver er tilgængelige fra forskellige danske biobanker samt SSI. Se svar fra de forskellige informantinterviews i Appendiks C.

Delkonklusion

Alle relevante data om patienten noteres i Sygehusafdelingernes PAS, som herefter indberettes til forskellige registre i SST, der har hver deres fokus. Dette medfører at alle data er tilgængelige fra SST med undtagelse af data om familieforhold, der er tilgængelige fra SSI's egne registre.

4.6 Konklusion

Ud fra ovenstående kan de operationelle spørgsmål opstillet i afsnit 3.1 og dermed det initierende problem besvares.

Et dansk initiativ til en national biobank er taget af SSI, der har udarbejdet visioner for projektets fysiske arkitektur samt indhold. De sundhedsdata der skal indgå i biobanken skal flyttes eller kopieres fra SST's registre og registreres i et Biobankregister, mens de biologiske prøver skal arkiveres i en prøvebank. Funktionen af disse varetages af et Koordinerende center.

Status for den nationale biobank er, at arkitekturen er fastlagt, men der arbejdes dog indledningsvist kun med opbygningen af Prøvebanken, mens opbygningen af Biobankregisteret udskydes til Prøvebanken er på plads. Dog forventes det, at Biobankregisteret skal indeholde sundhedsdata fra SST's registre.

Videnskabsministeriet, Novo Nordisk Fonden og SST har alle kendskab til projektet. Med hensyn til dataerhvervelsen mener SST, at de har ekspertviden om deres registre samt indholdet i disse, og derfor er der ikke andre der kan varetage deres arbejde. Dette belyser modsatrettede holdninger mellem SSI og SST angående indhentning af data, som kan forhindre, at Biobankregisteret kommer til at fungere efter visionerne.

Den nationale biobank skal besidde følgende relevante sundhedsdata, der skal være tilgængelige for forskere: CPR-nummer, familieforhold, fødselstilstand, medicinske undersøgelser, behandling, diagnose, dødsårsag, patologisvar, samt data om blod- og vævsprøver.

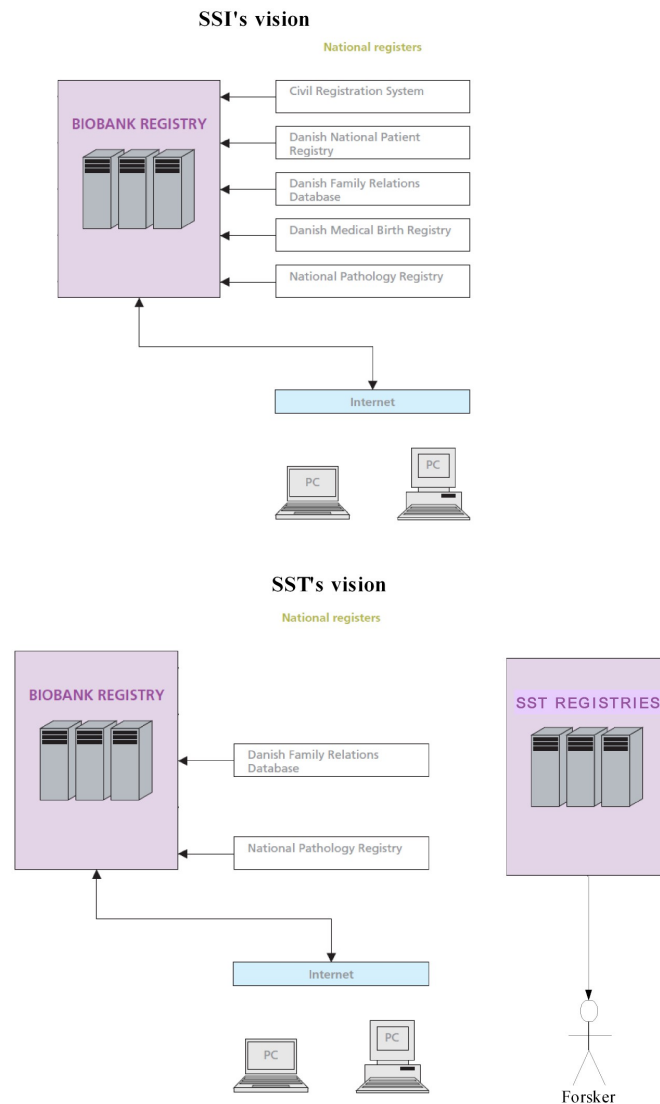
De relevante sundhedsdata er tilgængelige gennem SST's registre heriblandt Landspatient-

registeret og Patologiregisteret, samt SSI's egen Familie-relationsdatabase.

Overordnet kan det konkluderes at visionerne for det nationale biobanksprojekt er fastlagt, de relevante data er identificeret og ligeledes tilgængelige. På baggrund af dette anses etableringen af en national biobank i Danmark som værende praktisk mulig.

Problematisering

En vigtig del af konceptet for den nationale biobank består imidlertid i at have relevante sundhedsdata fra SST's registre, således at forskere let kan få adgang til både data og biologiske prøver. Der er dog en mulig problemstilling vedrørende datahåndteringen, idet SST ikke er interesseret i at videregive relevante sundhedsdata, da de ikke mener at andre instanser har kapacitet og viden om at håndtere dataregistreringen af sundhedsdata. Data kan som beskrevet ovenfor erhverves ved at kopiere eller flytte data fra SST til SSI, og det kan formodes at SST's holdning til datahåndteringen er tæt knyttet til valg af dataerhvervelsesmetode. Da SST besidder den overvejende del af de relevante sundhedsdata, er det essentielt, at SST vil indgå i et samarbejde om datahåndteringen i den nationale biobank, da formålet med denne ellers ikke kan opfyldes. Figur 4.3 på næste side illustrerer SSI's og SST's modsatrettede visioner for Biobankregisteret, som er udledt af projektgruppen på baggrund af interviews.



Figur 4.3: Skitsering af visioner for Biobankregisteret for hhv. SSI og SST udledt af projektgruppen på baggrund af de foretagne interviews men ikke verificeret af respondenterne.

4.7 Problemformulering

På baggrund af ovenstående fokuseres der på problemstillingerne vedrørende indhentning af sundhedsdata til den nationale biobank:

SSI ønsker at indhente data fra SST's registre, men SST ønsker ikke at dele data, og medmindre der findes en løsning på dette kan formålet med den nationale biobank ikke opnåes, idet den bygger på adgang til disse data.

Løsningsstrategi

Problemformuleringen i forrige afsnit konkretiserer problemstillingen vedrørende datahåndteringen for den nationale biobank, hvor der er uenighed mellem SST og SSI om placeringen af de relevante sundhedsdata. Da visionen for den nationale biobank er, at forskere kan tilgå relevante sundhedsdata samt biologiske data og prøver fra samme instans, er det nødvendigt, at parterne finder et fælles udgangspunkt for datahåndteringen, hvis visionerne skal indfries. Er dette ikke muligt, vil den nationale biobank ikke have nogen gavnvirkning i forhold til de muligheder forskere har i dag for at tilgå sundhedsdata og biologiske prøver, og en etablering af denne vil derfor være irrelevant. Derfor ønskes det at skitsere en løsning på problemstillingen beskrevet i forrige afsnit, ved at identificere et kompromis og skitsere en teknisk løsning, der understøtter dette kompromis. Den tekniske løsning skitseres vha. datahåndteringsarkitektur samt tekniske metoder til udførelsen af denne, og med inddragelse af forskerens snitflade. Formålet med skitseringen af datahåndteringsarkitekturen er at give et konkret bud på, hvordan det valgte kompromis bedst kan understøttes. Derfor evalueres det, hvorvidt kompromiset, datahåndteringsarkitekturen og de tekniske metoder er mulige at udføre. Ydermere undersøges arbejdsdelings- og økonomiske konsekvenser for både SST og SSI, ved udførelse af kompromiset, da disse kan have betydning for om det ønskes at udføre projektet.

Et kompromis kan bestå i, at SST ikke får noget valg i forhold til udlevering af data og dermed pålægges at udlevere dem til SSI. Dermed vil Biobankregisteret opbevare alle relevante sundhedsdata, og SST vil derfor ikke længere have adgang til disse. Fordelen ved dette kompromis er, at alle relevante sundhedsdata er placeret i Biobankregisteret, og der kan derfor ikke være tvivl om, hvor henvendelsen skal rettes, når data ønskes udtrukket. Ydermere er det muligt at kombinere stamdata om biologiske prøver samt de biologiske prøver fra Prøvebanken med de relevante sundhedsdata. Der er dog nogle ulemper forbundet med dette kompromis. Kompromiset bevirker, at SST mister data og dermed nogle af sine hidtidige funktioner. Dette skyldes, at der vil forekomme en reducere i antallet af ansøgninger om dataudtræk, idet forskere kan have brug for data, der befinder sig hos

SSI, og derfor retter henvendelse til dem. Yderligere vil de involverede registre ikke længere have samme værdigrundlag, idet nogle centrale registerdata er udtrukket til SSI. En anden ulempe er, at SSI ikke har nogen erfaring indenfor registerhåndtering samt viden om data. Et andet kompromis består i, at SST får ejerskabet over den nationale biobank og samtidig varetager driften af denne, således at sundhedsdata og biologiske prøver er samlet hos SST. Ligesom i det første kompromis er fordelene ved dette kompromis, at alle relevante sundhedsdata er placeret ét sted, i dette tilfælde hos SST, og der kan derfor ikke være tvivl om, hvor henvendelse om dataudtræk skal rettes. Ligeledes kan forskerne kombinere stamdata om biologiske prøver samt de biologiske prøver fra Prøvebanken med de relevante sundhedsdata. Ulempen ved at lade SST varetage driften af den nationale biobank er, at de ikke har erfaring indenfor håndteringen af biologiske prøver samt etablering af en national biobank. Yderligere mister initiativtagerne til projektet, SSI, ansvaret for den nationale biobank.

Et tredje kompromis indebærer en arbejdsdeling mellem de to parter, hvor SST beholder funktionen som registeransvarlig og derfor besidder de relevante sundhedsdata, samtidig med at SSI ligeledes har adgang til en kopi af de relevante sundhedsdata fra Biobankregisteret. Fordelen ved dette kompromis er, at SST ikke mister sine registerdata, og de vil derfor stadig være registeransvarlige, samtidig med at SSI har ansvaret for den nationale biobank. Endvidere giver dette kompromis også mulighed for, at forskere kan tilgå stamdata om biologiske prøver samt de biologiske prøver fra Prøvebanken og de relevante sundhedsdata i samme arbejdsgang. Ønsker forskere ikke indsigt i biologiske prøver, kan de rette henvendelse til SST's Forskerservice, der vil behandle deres forespørgsel på samme vis som de gør det i dag. Fordelen ved dette er, at SST kun vil miste et fåtal af funktioner. Ulemperne er dog, at SST ikke kan undgå at miste nogle af sine funktioner, og samtidig kan SST og SSI begge ansøges om dataudtræk, hvilket kan forvirre forskeren.

Ved at tage udgangspunkt i de beskrevne fordele og ulemper for hvert kompromis, kan det udledes, at hvert kompromis har forskellige konsekvenser for de involverede parter. Det første kompromis vil forventeligt medføre utilfredshed hos SST, idet de vil miste funktionen som registeransvarlig for nogle af registrene. Samtidig er det uklart, om SSI er gearet til at være ansvarlig for registreringen af de relevante sundhedsdata. Omvendt ville det andet kompromis formodentlig medføre utilfredshed hos SSI, da de er ideskaberne for den nationale biobank og derfor ikke har interesse i at give ideen videre. Samtidig kan det ikke forventes, at SST ønsker at stå for driften af en biobank, idet det ikke er et område de har erfaring indenfor. Det tredje kompromis er derimod den løsning, hvor der opnåes mindst utilfredshed hos begge parter, og hvor SST bibeholder størstedelen af deres funktioner. Det har ikke været muligt at få kendskab til, hvorvidt dette arbejdsdelingskompromis var SSI's hensigt fra begyndelsen, da dette ikke blev klarlagt ud fra interviews med Mads Melbye og Kristian Hveem samt litteraturstudiet af visionerne for den nationale biobank. Ligeledes blev dette ikke klargjort gennem interview med SST som fastholdt, at de relevante

sundhedsdata skal blive hos dem. Hvis ovenstående kompromis har været SSI's hensigt i forhold til dataerhvervelse, tyder uoverensstemmelserne mellem SST og SSI på, at SSI har manglet at tydeliggøre argumenterne for, hvorfor dette kompromis er den bedste løsning.

Ud fra ovenstående fremgår det, at det tredje kompromis bedst udnytter de nuværende kompetencer hos både SSI og SST som ansvarlige for hhv. biologiske prøver og sundhedsdata, og vigtigst af alt ændrer det tredje kompromis mindst muligt på den nuværende arbejdsdeling. På denne baggrund vurderes det, at det tredje kompromis må være den bedste løsning.

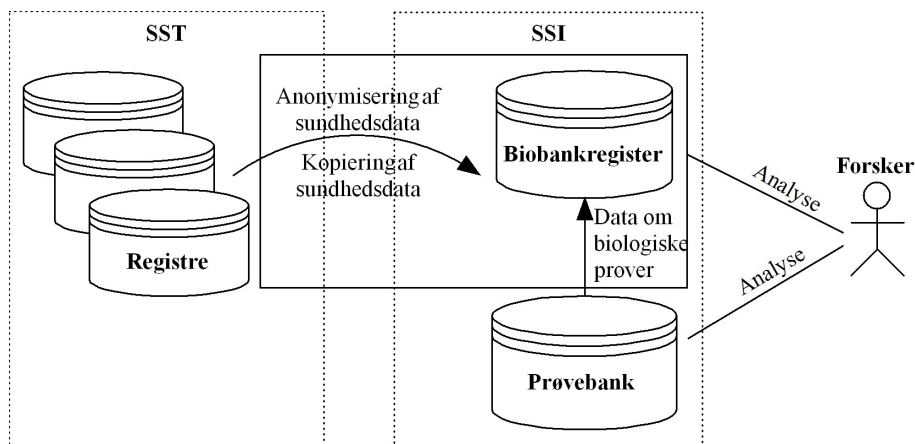
Løsningsstrategien er derfor at skitsere arkitekturen for et system, der understøtter hele datahåndteringsprocessen og dermed det valgte kompromis. Dette gøres via udarbejdelse af visioner, krav, analyse, design for systemet samt evaluering af hvorvidt det er muligt at udføre kompromiset, datahåndteringsarkitekturen og de tekniske metoder. For at vurdere, hvorvidt dette kompromis er hensigtsmæssig at indføre, vil der foretages en konsekvensvurdering af kompromisets betydning for SST og SSI.

Vision for datahåndteringssystemet

I løsningsstrategien blev det valgt at tilgå problemstillingen ved at identificere et kompromis der sikrer, at SST bibeholder deres funktion som registeransvarlig, samtidig med at SSI har adgang til de relevante sundhedsdata fra Biobankregisteret. Visionen er derfor at foretage datahåndteringen vha. tekniske standardmetoder. Derfor skal visionerne for den nationale biobank tages i betragtning, idet kompromiset så vidt muligt skal forsøge at tilgodese disse. Som beskrevet i foranalysen er visionen for den nationale biobank at lave et Biobankregister, hvorfra forskere via en online webportal kan få et unikt indblik i aggregerede sundhedsdata samt antallet af patienter, for hvem der eksisterer biologiske prøver. Ydermere skal forskere ved kontakt til SSI have mulighed for at få adgang til rå sundhedsdata og biologiske prøver samt stamdata om disse.

Udgangspunktet for kompromiset er, at SST og SSI finder et fælles udgangspunkt for datahåndteringen i den nationale biobank. Datahåndteringen skal foretages vha. standardmetoder, da disse er afprøvede og velkendte i litteraturen. For at SST skal beholde deres registerdata, samtidig med at Biobankregisteret hos SSI skal indeholde relevante sundhedsdata, er det nødvendigt at kopiere de relevante sundhedsdata fra SST til Biobankregisteret. Dette scenarie for dataerhvervelsen er lig scenariet, hvor SST i dag udtrækker indberettede data fra en registerdatabase til deres data warehouse, se Appendiks C. På baggrund af dette vurderes det, at den samme fremgangsmåde kan benyttes af SSI til erhvervelse af data fra SST. For at forskere kan få udtræk af relevante sundhedsdata samt biologiske prøver og stamdata om disse, er det nødvendigt at anonymisere de relevante sundhedsdata, hvilket kan foretages ved at tildele et ID til alle CPR-numre, som er forskelligt fra CPR-nummeret. Da Biobankregisteret skal besidde og samtidig modtage de relevante sundhedsdata fra flere registre hos SST, skal det kunne håndtere store datamængder samt lagring af disse i det samme format, hvormed de lagres hos SST. Ligeledes bør Biobankregisteret indeholde de nyeste sundhedsdata, og det skal derfor opdateres med et prædefineret tidsinterval, hvilket anses for at være muligt, idet SST's data warehouse automatisk opdateres en gang i døgnet. Da forskere skal have mulighed for at koble relevante sundhedsdata samt stamdata om bio-

logiske prøver, skal denne funktionalitet håndteres. Datahåndteringen skal foregå mellem SST og SSI og figur 6.1 illustrerer datahåndteringssystemets grænser samt dataflow.



Figur 6.1: Datahåndteringssystemets grænser, visualiseret ved den sorte kasse, samt dataflow for den nationale biobank.

Visionen for datahåndteringssystemet er derfor at SST besidder de relevante sundhedsdata og SSI udelukkende kopierer de relevante sundhedsdata fra SST til Biobankregisteret. For at konstruere dette datahåndteringssystem skal systemet benytte sig af tekniske standardmetoder til at håndtere kopiering og anonymisering af de relevante sundhedsdata, store datamængder, opdatering af Biobankregisteret samt kobling mellem data om de biologiske prøver og sundhedsdata i Biobankregisteret. Disse metoder vil derfor identificeres i de følgende kapitler.

Metode til systemudvikling

For at skitsere og designe datahåndteringsarkitekturen samt standardmetoderne til datahåndteringssystemet er det fordelagtigt at anvende en systemudviklingsmetode til styring af udviklingsprocessen. Dette projekt fokuserer dog udelukkende på en konceptudvikling, idet de identificerede standardmetoder kun designes og dermed ikke implementeres, hvorfor systemudviklingsmetoden kun vil være inspirerende. I dette kapitel vil valg af systemudviklingsmetode blive beskrevet efterfulgt af en beskrivelse af, hvordan denne kan bruges som inspiration i dette projekt.

7.1 Valg af systemudviklingsmetode

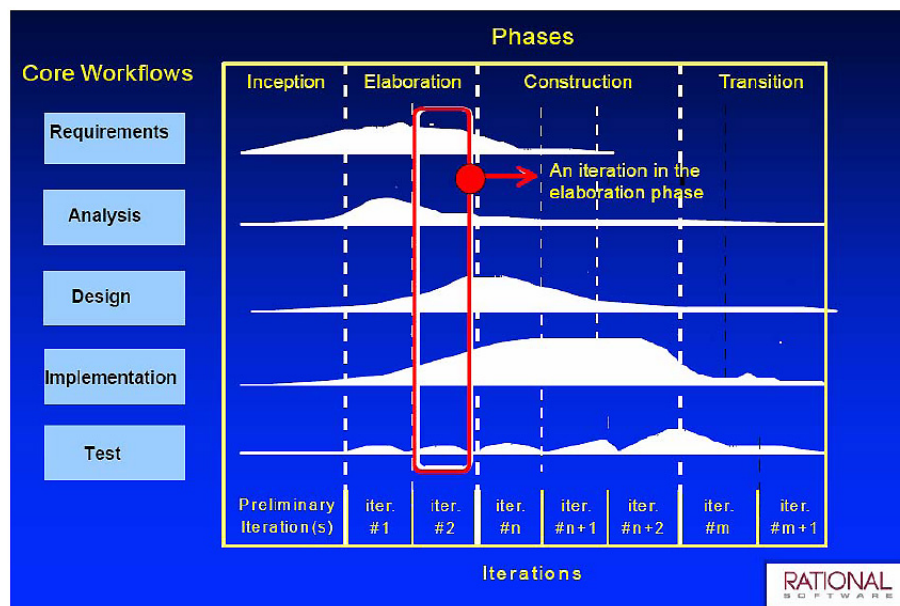
Til styring af systemudviklingsprocessen kan der anvendes forskellige systemudviklingsmetoder såsom Vandfalds-modellen, V-modellen, Spiral-modellen og Unified Process (UP).

Vandfalds-modellen består af forskellige workflows og illustrerer de forventelige hændelser samt deres rækkefølge. Påbegyndelsen af et nyt workflow sker først når det tidligere workflow er helt færdiggjort, og der er ikke mulighed for at vende tilbage til et tidligere workflow. V-modellen er en variation af vandfalds-modellen, hvor det demonstreres, hvordan testaktiviteterne er forbundet med Analyse og Design. Viser testen problemer med systemet er det muligt at gå tilbage til Analyse, Design og kodning og rette disse, hvorefter en ny test udføres. Spiral-modellen kombinerer udviklingsaktiviteter med risikohåndtering for at minimere og kontrollere risici. Udviklingen foregår i iterationer, hvor der efter hver iteration tages stilling til risikohåndtering. UP er en iterativ og inkrementel metode, der inddeles i følgende fire faser: Forberedelse, Etablering, Konstruktion og Overdragelse, som alle udgøres af fem work flows: Krav, Analyse, Design, Implementering og Test. De enkelte faser deles op i en række iterationer, der inkluderer de fem work flows i højere eller mindre grad afhængig af antallet af iterationer. (Scott, 2001; Pfleeger, 2001)

I dette projekt vælges det at hente inspiration fra UP, da det vurderes at være den metode,

der er mest fleksibel i fht. tilføjelse af nye ideer samt revurdering af opstillede krav.

De fire faser samt fem workflows i UP er illustreret i figur 7.1.



Figur 7.1: De fire faser og de fem workflows i Unified Process.

I Appendiks D kan der opnåes yderligere information om UP, heriblandt de fire processer og fem workflows.

7.2 Anvendelse af Unified Process

Da dette projekt omhandler identificering af standardmetoder og dermed konceptudviklingen af et datahåndteringssystem anvendes følgende fire workflows: Krav, Analyse, Design og Test, idet disse er relevante i udviklingen af et systemkoncept. Faserne hvori disse workflows udføres er Forberedelse og Etablering, jævnfør figur 7.1.

Krav

Til opstilling af krav vil viden om systemets grænser og dataflow, opstillet i forrige kapitel, anvendes sammen med systemindhold, use case diagrammer, aktivitetsdiagrammer og deploymentdiagrammer. Krav opstilles hovedsageligt i Forberedelsesfasen, men yderligere krav kan findes i Etableringsfasen, idet nye krav kan opstå under Analyse og Design. Kravspecifikationer skal være en beskrivelse af, hvad der ønskes af et system og relateret til identificerede behov eller muligheder. Der findes både funktionelle og ikke-funktionelle krav. Funktionelle krav udledes fra use cases og beskriver interaktion mellem systemet og dets omgivelser, og identificerer de nødvendige opgaver, handlinger eller aktiviteter, der skal fuldføres. Ikke-funktionelle krav specificerer kriterier eller begrænsninger for løsningen

samt den måde systemet kan opbygges. (Scott, 2001)

Analyse

Analyse foretages i både Forberedelsesfasen og Etableringsfasen og formålet er at opnå en reel forståelse af kravene og bruge denne forståelse til at opnå viden om de identificerede standardmetoder som kan bruges videre i Design. (Scott, 2001) Analysen i dette projekt udføres dog anderledes sammenlignet med en Analyse til systemudvikling. Dette skyldes, at systemet skal udvikles på et konceptuelt niveau. Derfor vil Analysen fokusere på de standardmetoder systemets koncept skal bygge på, fremfor en udledning af f.eks. analyseklasser, som normalt er hele fundamentet for Design.

Design

Design foretages også i lille grad i Forberedelsesfasen, men er primært en del af Etableringsfasen. Det er en kreativ og iterativ proces, hvor de identificerede standardmetoder fra Analysen transformeres til løsninger, som kan opdeles i et konceptuel og teknisk design. Design i dette projekt tager udgangspunkt i Analysen, ved at fokusere på specifikke designmuligheder for hver af de identificerede standardmetoder til systemets udvikling. Ligesom i Analysen er Design også forskellig fra hvad der normalt forventes, da der ikke vil være fokus på udledning af designklasser og interaktionsdiagrammer.

Test

Formålet med Test er at sikre, at systemet opfylder de funktionelle og ikke-funktionelle krav der opstilles i Krav samt den opstillede vision fra forrige afsnit. Normalt udføres systemtests som blackboks-tests udledt direkte fra use cases, hvilket ikke er muligt i dette projekt da det kun omhandler systemkonceptet for datahåndteringssystemet, hvorfor der foretages en evaluering af datahåndteringssystemets koncept. Dette gøres ved at bede relevante respondenter om deres vurdering af konceptet, hvorfor koncepttesten er subjektiv.

Del II

KRAV

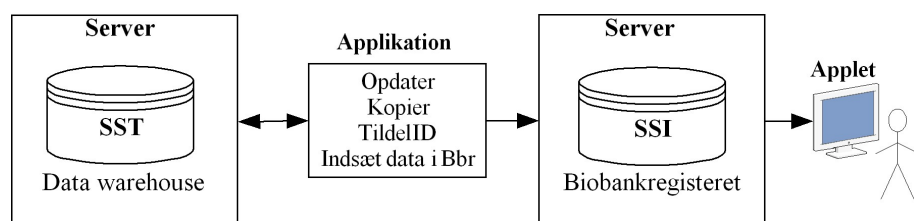
Formålet med denne del er at beskrive kravene til datahåndteringssystemets funktionalitet gennem systemindhold, use cases, aktivitetsdiagrammer med swimlanes, kravspecifikationer og et deploymentdiagram.

Krav

Krav til datahåndteringssystemet kan identificeres ud fra systemets indhold, use cases, aktivitetsdiagrammer med swimlanes samt deploymentdiagram.

8.1 Systemindhold

For at opnå adgang til en database der ligger på en server, skal der oprettes forbindelse til den server, hvorpå databasen ligger. SST's database ligger på en server hos SST, mens SSI's Biobankregister, skal ligge på SSI's server. Der skal derfor identificeres en struktur til kommunikation mellem SST's server og SSI's server, for at indhente relevante sundhedsdata til Biobankregisteret. Derfor er det oplagt, at datatransmissionen skal foregå mellem SST og SSI gennem et eksternt programmel, idet sådanne opgaver ofte løses af en applikation. Denne applikation skal ligeledes håndtere opdatering ved et prædefineret tidsinterval, tildeling af ID samt lagring. Ligeledes skal der være en applet, der tillader kommunikation mellem forskeren og Biobankregisteret. Med udgangspunkt i figur 6.1 kan systemindholdet, bestående af applikationen, Biobankregisteret og en applet, samt dataflow mellem SST og SSI illustreres ved hjælp af nedenstående figur.

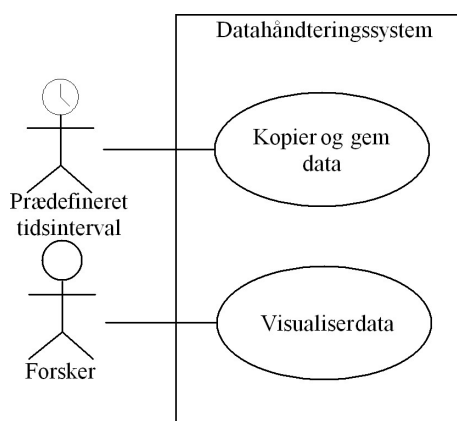


Figur 8.1: Systemindhold og dataflow mellem data warehouset hos SST, applikationen, Biobankregisteret hos SSI og forskeren. Pile begge veje visualiserer, at forespørgsler foretages til databasen med efterfølgende datatransport fra databasen til applikationen. Pile én vej illustrerer, at data kommunikeres fra applikationen til Biobankregisteret og fra Biobankregisteret til forskeren.

8.2 Use Cases

Use cases udtrykker en brugssituation givet af et system. Use case diagrammer visualiserer use cases og afbilder et antal eksterne aktører samt deres forbindelse til use cases. (Eriksson et al., 2004a)

På baggrund af visionerne for Biobankregisteret beskrevet i afsnit 6 kan følgende to use cases udledes: UC: "Kopier og gem data" samt UC: "Visualiser data". Aktørerne til disse use cases er et prædefineret tidsinterval samt forskere. Nedenstående figur 8.2 illustrerer disse i et use case diagram.



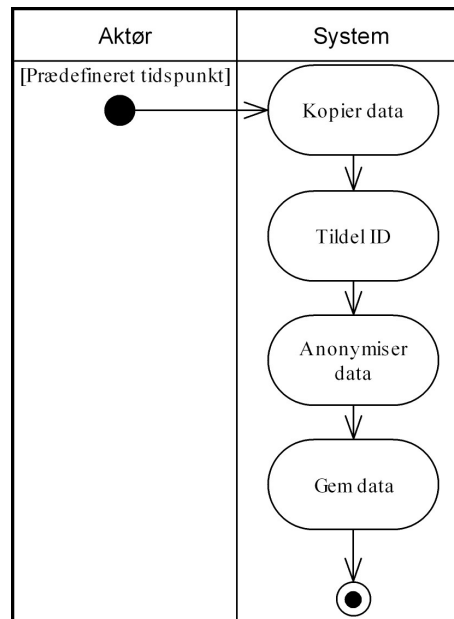
Figur 8.2: Use case diagram for datahåndteringssystemet med visualisering af de forskellige aktører og use cases.

8.3 Aktivitetsdiagrammer med swim lanes

Aktivitetsdiagrammer består af handler der udgør en aktivitet, og illustrerer derfor et sekventielt flow af aktiviteter udført i use casen. Ved anvendelse af swim lanes i et aktivitetsdiagram illustreres aktøren til hver eneste handling. For at visualisere konceptet ved hver use case er to swim lanes defineret: aktør og system. (Eriksson et al., 2004b)

UC: "Kopier og gem data"

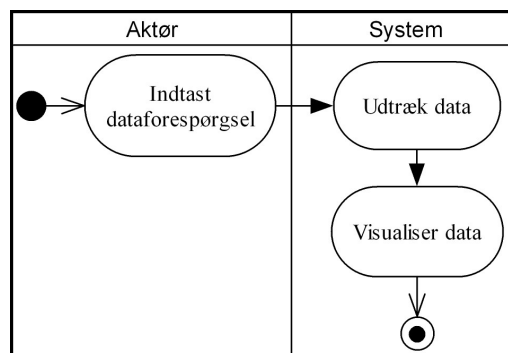
Dette er brugssituationen, hvor Biobankregisteret bliver opdateret med nye sundhedsdata. Selve processen initieres på et prædefineret tidspunkt, hvor relevante sundhedsdata udtrækkes fra SST's registre. De identificerede data kopieres og gøres klar til anonymisering ved at tildele hvert CPR-nummer et ID, hvorefter de kopierede data samt ID lagres i Biobankregisteret, jævnfør figur 8.3 på næste side.



Figur 8.3: Aktivitetsdiagram med swim lanes for UC: "Kopier og gem data".

UC: "Visualiser data"

Dette er den brugssituation, hvor forskeren ønsker at visualisere data fra Biobankregisteret. Forskeren initierer processen ved at indtaste en dataforespørgsel, hvorefter systemet udtrækker de ønskede sundhedsdata og visualiserer disse for forskeren, se figur 8.4.



Figur 8.4: Aktivitetsdiagram med swim lanes for UC: "Visualiser data".

8.4 Kravspecifikation

På baggrund af visionerne beskrevet i kapitel 6, ovenstående use cases, aktivitetsdiagrammer med swim lanes og viden om datasikkerhed kan følgende funktionelle og ikke-funktionelle krav opstilles:

Funktionelle krav:

- Opdatering af Biobankregisteret skal forekomme ved et prædefineret tidsinterval

- Relevante sundhedsdata skal kopieres fra SST
- CPR-numre skal tildeles et ID
- Relevante sundhedsdata skal anonymiseres
- Relevante sundhedsdata skal lagres hos SSI
- Relevante sundhedsdata skal visualiseres for forskeren

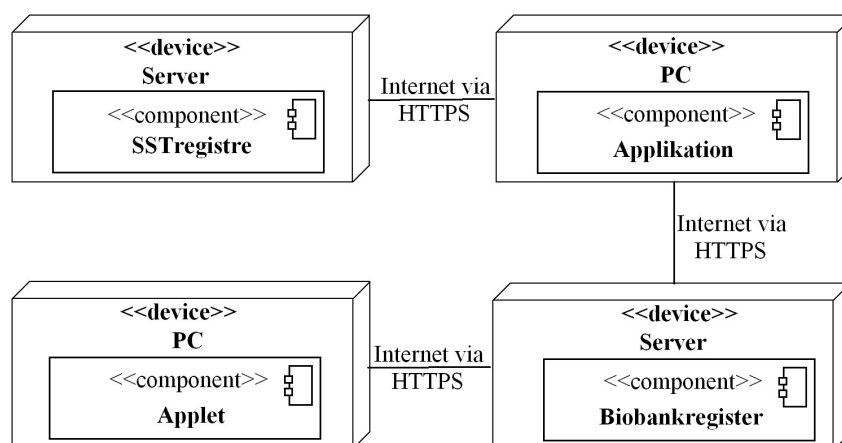
Relevante sundhedsdata henviser til de data der blev identificeret i Foranalysen som relevante for den nationale biobank.

Ikke-funktionelle krav:

- Store datamængder skal håndteres
- Stamdata om biologiske prøver og sundhedsdata skal kobles
- Datasikkerhed skal håndteres
- Relevante sundhedsdata skal forblive i samme format som hos SST (se Appendiks E for viden om de anvendte formater)

8.5 Deploymentdiagram

Deploymentdiagrammer illustrerer og beskriver strukturen af systemets hardwareenheder og den software der eksekveres på hver enhed. I dette system er koden eksekveret på servere og to applikationer, hvilket udgør systemets hardware. Applikationerne kommunikerer med serverne samt forskeren over Internettet via HTTPS, se figur 8.5. (Eriksson et al., 2004c)



Figur 8.5: Deploymentdiagram for systemet indeholdende de to servere hos hhv. SST og SSI samt applikationerne til kommunikation mellem de to servere og forskeren og Biobankregisteret.

Del III

ANALYSE

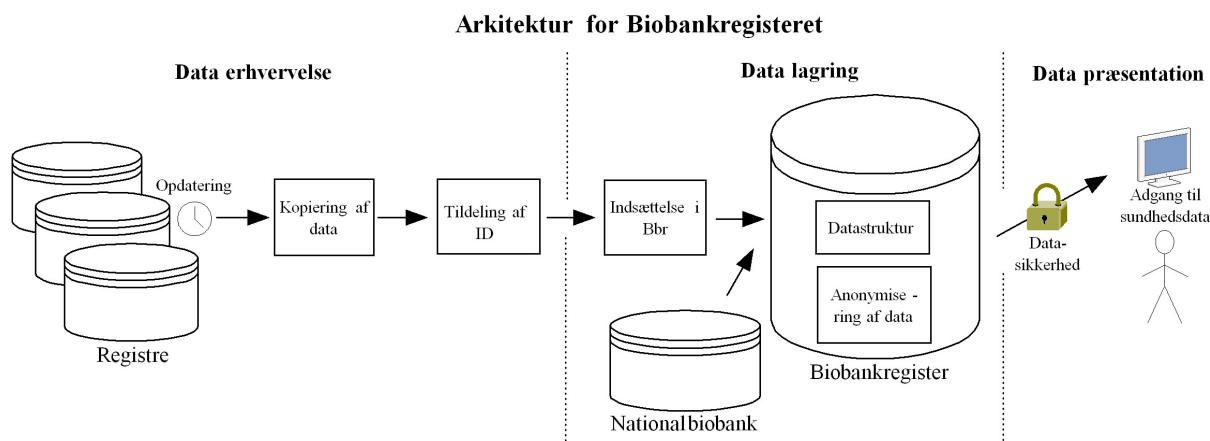
Analysen omhandler metoderne til at indfri de funktionelle og ikke-funktionelle krav, og dermed udfærdigelse af datahåndterignssystemet med henblik på at dette kan designes.

Systemets dele

Dette kapitel fokuserer på systemarkitekturen for datahåndteringen og dermed datahåndteringssystemet samt understøttelse af systemarkitekturen.

9.1 Systemarkitektur

I kapitel 5 blev Løsningsstrategien specificeret, og det blev valgt at skitsere arkitekturen for et system, der understøtter datahåndteringen mellem SST og SSI. For at udlede systemarkitekturen i forbindelse med datahåndteringen tages der derfor udgangspunkt i figur 6.1 på side 40, idet denne skitserer flow af data fra SST til SSI, samt systemindhold i afsnit 8.1 på side 47, da dette visualiserer de anvendte standardmetoder. Ud fra disse figurer kan det udledes at datahåndteringen kan nedbrydes i dataerhvervelse, datalagring og datapræsentation, som er de typiske processer der varetages af et datawarehouse. Yderligere kan de identificerede standardmetoder, som applikationen skal foretage, blive illustreret i den rækkefølge de skal foretages. Indledningsvist skal opdateringsprocessen initieres, hvorefter kopieringen af relevante sundhedsdata og tildelingen af ID skal foregå mhp. indsættelse i Biobankregisteret. Efterfølgende skal disse lagres og anonymiseres i Biobankregisteret. Datatransformation illustreres ikke, da det er et ikke-funktionelt krav, at de relevante sundhedsdata skal bevare samme format som hos SST. Datahåndteringssystemets systemarkitektur kan illustreres ud fra figur 9.1 på næste side, der overordnet illustrerer de tre processer samt de .



Figur 9.1: Systemarkitekturen for datahåndteringssystemet med de identificerede standardmetoder.

Løsningsstrategien var at skitsere arkitekturen for et system til understøttelse af kompromiset, hvorfor ovenstående figur vil være styrende for de følgende kapitler samt Design, som derfor vil fokusere på processerne dataerhvervelse, datalagring samt datapræsentation. Under hvert kapitel vil de tilhørende identificerede standardmetoder blive beskrevet.

Understøttelse til systemarkitektur

Datawarehouse

Da formålet med Biobankregisteret er at samle store datamængder på et sted, hvorfra data let kan udtrækkes og analyseres, kan Biobankregisteret med fordel være indeholdt i et data warehouse. Selve grundlaget for et data warehouse er en database, der indeholder elektronisk indkomne data fra forskellige kilder, organiseret således at rapporter og analyse af data kan foretages bedst muligt. Den overordnede struktur for, hvordan data processeres, lagres og anvendes i data warehouset kan beskrives som data warehousets dataarkitektur. Dataarkitektur beskriver hvordan erhvervelse, lagring og præsentation af data foregår og er relateret, og er derfor central i forhold til udviklingen af Biobankregisteret. Regler, procedurer og funktioner, der tillader data warehouset at arbejde og opfylde kravene, er ligeledes dataarkitektur. (Ponniah, 2001; Beynon-Davies, 2000a)

De tre centrale processer, der varetages af et data warehouse, kan beskrives således:

- **Data erhvervelse:** styrer rengøringen, transformationen og integrationen af data.
- **Data lagring:** inkluderer læsningen af data fra dataerhvervelsen, og ligeledes lagringen af data i passende formater med henblik på præsentation af data.
- **Data præsentation:** inkluderer afhængige datacentre, specielle multidimensionelle databaser, og en lang række af forespørgsels- samt rapporteringsmuligheder.

(Ponniah, 2001)

Database Management System

Dataarkitekturen for Biobankregisteret blev klarlagt i ovenstående afsnit. For at lette håndteringen med lagring og udtræk af data fra databasen er næste trin at udvælge en datamodel og dermed et DBMS til håndtering af Biobankregisteret, der skal besidde egenskaber til at håndtere store datamængder.

Den valgte datamodel specificerer den logiske organisering af databasen samt, hvorledes adgang til information i databasen kan opnås. Et DBMS er et softwareprogram, der kontrollerer oprettelsen, vedligeholdelsen samt brugen af en database med en computer som platform. Yderligere giver et DBMS mulighed for at kontrollere dataadgang, sikre dataintegritet, håndtere samtidige brugerforespørgsler og gendanne databaser. (Elmasri og Navathe, 2003a) Nogle af fordelene ved at bruge et DBMS er listet i det følgende:

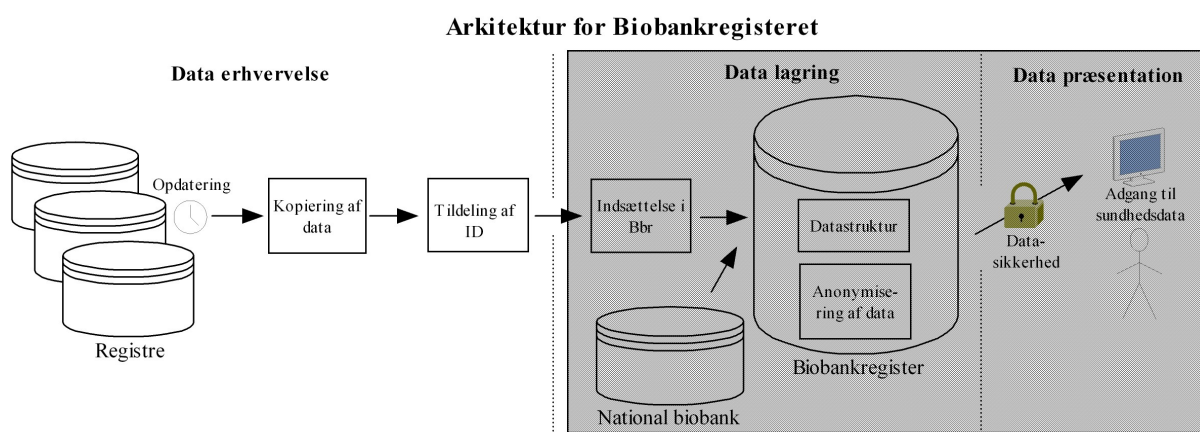
- **Kontrollering af redundans:** Udføres så hvert logisk data element gemmes ét sted i databasen med henblik på at sikre konsistens og gemme lagringsplads til andet data. Dog kan det i praksis nogle gange være nødvendigt at kontrollere redundans for at forbedre udførelsen af forespørgsler.
- **Begrænsning af adgang:** Når flere brugere skal dele en database er det normalt at brugerne ikke har autorisation til at tilgå al information i databasen. Dette gælder specielt ved personhenførbare data, som kun autoriserede brugere har adgang til.
- **Sikring af lagringsstrukturer til effektiv forespørgselsprocessering:** Forespørgselsprocesseringen og optimeringsmodulet i et DBMS er ansvarlig for valget af en effektiv forespørgselsudførselsplan for hver forespørgsel baseret på de eksisterende lagringsstrukturer.

(Elmasri og Navathe, 2003a)

Der findes forskellige datamodeller til styring af et DBMS, herunder bl.a. relationelle og objektorienterede. Til håndtering af Biobankregisteret vælges det at anvende en relationel datamodel, da det er en udbredt simpel model til håndtering af store datamængder, der repræsenterer databasen som et delt opbevaringssted udgjort af flere tabeller. Den relationelle model repræsenterer databasen som en samling af relationer, hvor hver relation ligner en tabel af værdier. Når en relation er tænkt på som en tabel af værdier skyldes det, at hver række i tabellen repræsenterer en samling af relaterede dataværdier. Relationelle DBMS anvendes ofte, hvis det ønskes at gruppere data i forskellige kategorier. Det relationelle DBMS understøtter kun bestemte datatyper såsom integers, tidsangivelser, strings osv., hvilket er brugbart i traditionelle anvendelsesområder såsom administrativ dataprocessering. Standardsproget for oprettelse, manipulering og forespørgsel af data i et relationel DBMS er Structured Query Language (SQL). (Elmasri og Navathe, 2003a)

Kapitel 10

Dataerhvervelse



Figur 10.1: Med udgangspunkt i figur 9.1 kan dataerhvervelsesprocessen for datahåndteringssystemet illustreres.

Som illustreret i ovenstående figur omhandler dataerhvervelsesprocessen kopiering af relevante sundhedsdata fra de forskellige registre hos SST, transport af de kopierede data til et samlet område og derefter tildeling af ID til CPR-nummeret med henblik på lagring i Biobankregisteret.

10.1 Tidsinterval for opdatering

Ved etablering af Biobankregisteret skal der tages højde for tidsintervallet for opdatering af databasen, således at forskere har adgang til nyligt indberettede sundhedsdata. Hos SST opdateres Landspatientregisteret hver nat med de nyeste indberettede sundhedsdata, se Appendiks C. Der skal derfor tages stilling til, hvor ofte Biobankregisteret skal opdateres, hvorfor et prædefineret tidsinterval, der sikrer jævnlige opdateringer af Biobankregisteret, skal fastsættes. Da det er valgt at udtrække alle relevante sundhedsdata, nye som gamle, skal Biobankregisteret ryddes for data inden de nye sundhedsdata lagres, hvorfor opdaterin-

gen må være en krævende proces (Elmasri og Navathe, 2003b). Ydermere er det primære formål med Biobankregisteret, at det skal anvendes til forskning, og derfor er det ikke overvejende kritisk, at de nyeste sundhedsdata ikke er tilgængelige i Biobankregisteret fra dag til dag. På baggrund af dette, vælges det at lave en ugentlig opdatering af Biobankregisteret, hvor såvel relevante sundhedsdata som stamdata om biologiske prøver opdateres.

10.2 Kopiering af relevante data

Ifølge Visionen for datahåndtering er det klart, at kopiering af relevante sundhedsdata fra SST var det centrale element i kompromiset, idet kun de relevante sundhedsdata kopieres, og SST dermed mister et minimum af deres funktioner.

For at kopiere de relevante sundhedsdata skal det identificeres, hvilke registre der besidder disse data. I foranalysen blev følgende relevante sundhedsdata identificeret som de centrale sundhedsdata, der skal være indeholdt i Biobankregisteret: CPR-nummer, familieforhold, fødselstilstand, medicinske undersøgelser, behandling, diagnose, dødsårsag, stamdata om biologiske prøver og patologisvar, som findes i forskellige registre hos SST og SSI. Følgende registre skal derfor anvendes for at kopiere disse relevante sundhedsdata til Biobankregisteret:

- Landspatientregisteret: CPR-nummer, medicinske undersøgelser, diagnose og behandling
- Dødsårsagsregisteret: dødsårsag
- Fødselsregisteret: fødselstilstand
- Familie-relationsregisteret: familieforhold
- Patologiregisteret: patologisvar
- Prøvebanken: stamdata om biologiske prøver

Når Biobankregisteret skal opdateres, skal hele Biobankregisteret og dermed alle tabeller i Biobankregisteret overskrives med helt nye sundhedsdata. Dette betyder, at Biobankregisteret ikke kun bliver opdateret med nyeste sundhedsdata fra SST, men at alle tabeller ryddes for indhold og de relevante sundhedsdata, nye som gamle, lagres på ny når opdateringen af Biobankregisteret forekommer. Denne opdateringsmetode vælges for at mindske fejlforekomster under opdateringsprocessen, hvor nye sundhedsdata skal placeres korrekt i forskellige tabeller i Biobankregisteret.

Da kopieringen kun skal forekomme på de relevante sundhedsdata i de forskellige registre, laves en partiel kopiering (Elmasri og Navathe, 2003b). Til håndtering af kopieringsprocessen laves en applikation, der håndterer kommunikationen mellem SST's og SSI's DBMS samt kopieringen af de relevante sundhedsdata.

10.3 Tildeling af ID

Da det skal være muligt for forskere at ansøge om dataudtræk og biologiske prøver, skal personhenførbare data anonymiseres, da CPR-numre kun udleveres i sjældne tilfælde (Folketinget, 2010). Dette betyder, at hvert CPR-nummer skal tildeles et ID efter kopieringsprocessen, med henblik på udelukkende at anvende CPR-nummeret som fællesnævner for relevante sundhedsdata og biologiske prøver. Fællesnævneren er en vigtig parameter, idet det skal være muligt at koble relevante sundhedsdata fra Biobankregisteret med biologiske prøver fra Prøvebanken.

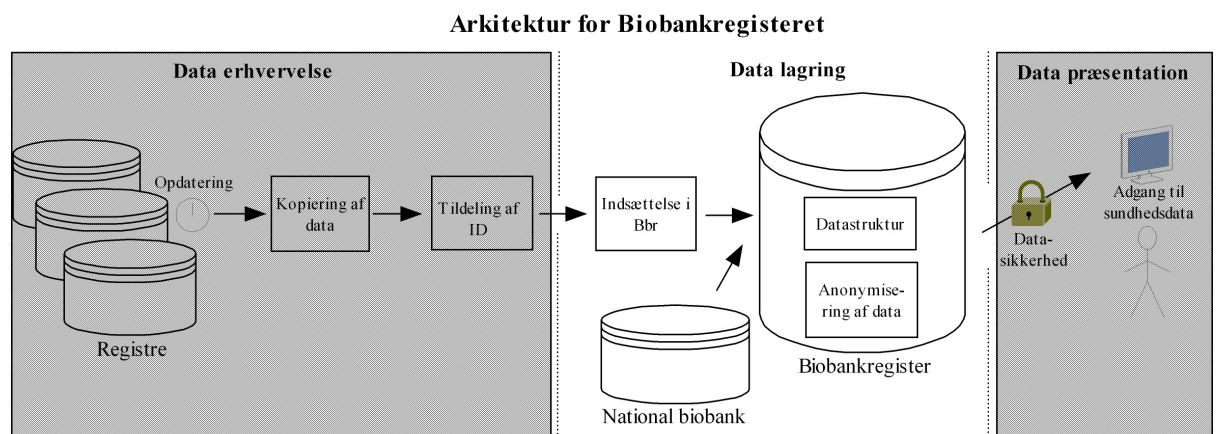
Identificeringen af ID kan foregå ved forskellige metoder:

- **Datasubstitution:** Udskifter data i en kolonne med anden lignende information, der ikke er relateret til de virkelige data (f.eks. erstatning af CPR-nummer med et andet ID). Datasubstitution er en effektiv metode i fht. at bevare udseendet af databasen. Dog skal der findes mange forskellige kombinationer af substitutioner.
- **Dataomrokering:** Minder om datasubstitution, men her findes de substituerende data ved at omrokere data i kolonnen. En ulempe ved omrokering er, at de rigtige data stadig kan findes for den enkelte person og dermed er der risiko for at data kan blive personhenførbare.
- **Varians:** Varians er brugbart på numerisk data og indebærer, at hver numerisk værdi i en kolonne modificeres ved en tilfældig procentdel af dens sande værdi. Ulempen ved denne metode er, at det f.eks. ikke er muligt at tilbageidentificere patienten.

(Net 2000 Ltd., 2005)

I dette projekt anvendes datasubstitution ved at tildele hver patient et ID, således at ID'et kobles til CPR-nummeret, og dermed er det der visualiseres for forskeren. Grunden til at CPR-nummeret ikke skal overskrives er, at det skal være muligt at koble relevante sundhedsdata med biologiske prøver. Derfor skal SSI tildele alle registrerede patienter et ID.

Datalagring



Figur 11.1: Med udgangspunkt i figur 9.1 kan lagringsprocessen for Biobankregisteret illustreres.

Datalagring dækker processen med at lagre de kopierede og anonymiserede sundhedsdata i Biobankregisteret, som illustreret i figur 11.1. Lagringen af sundhedsdata bør ske på den mest hensigtsmæssige måde så forskeres adgang til data lettes. Ydermere skal de relevante sundhedsdata anonymiseres.

11.1 Indsættelse af relevante data i Biobankregisteret

Når de relevante sundhedsdata er kopierede og alle CPR-numre er tildelt et ID og klar til lagring i Biobankregisteret. Indsættelse af de kopierede sundhedsdata i Biobankregisteret håndteres af Java applikationen.

11.2 Datastruktur

En af fordelene ved Biobankregisteret er, at det bliver muligt for forskerne at koble relevante sundhedsdata med stamdata om biologiske prøver. Som beskrevet i afsnittet om håndtering af store datamængder kan lagring af data håndteres af et DBMS. Til håndtering af Biobankregisteret blev det valgt at anvende et relationelt DBMS, hvorfor den konceptuelle datamodel kan afbildes med et Entitets/Relations-diagram (ER-diagram). ER-modellering specificerer, hvilke data der skal lagres, samt hvordan de er forbundet. Selve initieringen af datalagring kan håndteres af Java applikationen.

ER-diagram

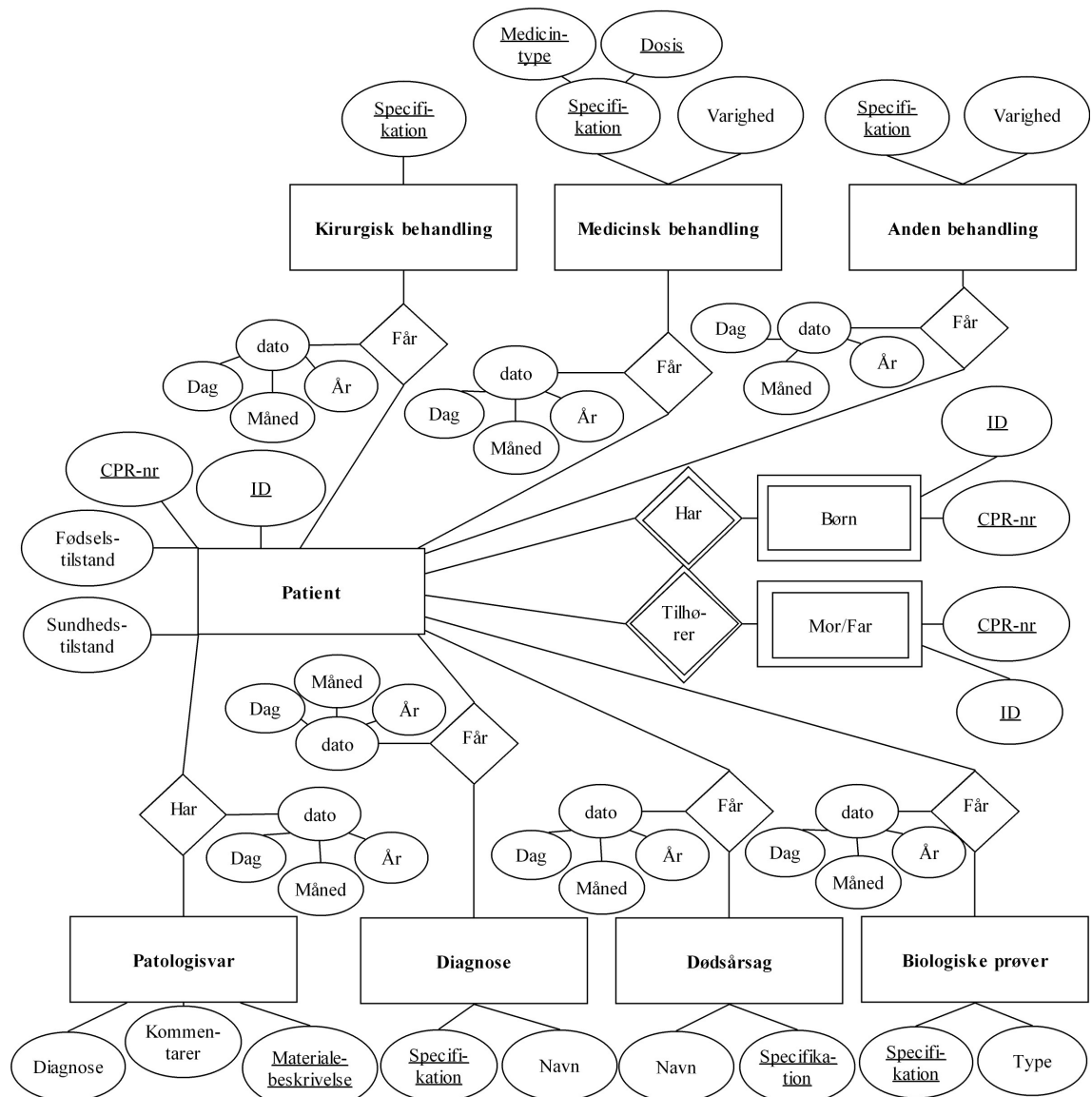
ER-diagrammer er baseret på en beskrivelse af en mini-verden bestående af entiteter samt relationer mellem entiteter. En entitet er en "ting" eller et "objekt" i mini-verdenen som kan skelnes fra andre objekter vha. nøgler, der unikt beskriver entiteten, og den kan være enten fysisk (fx. Patient) eller konceptuel (fx. Patologisvar). (Silberschatz et al., 2001a)

Et entitetssæt er en samling af flere entiteter, der deler de samme egenskaber (attributter). Entiteter karakteriseres ved hjælp af attributter, som beskriver de egenskaber hvert medlem af et entitetssæt besidder. Attributterne for et entitetssæt udtrykker, at databasen lagrer tilsvarende information for hver entitet i entitetssættet, men hver entitet kan have en forskellig værdi for hver attribut. Dermed har alle entiteter i et entitetssæt de samme attributter, men attributterne kan have forskellige værdier. Entiteter kan være enten stærke eller svage. For at en entitet skal være stærk, skal den have en primærnøgle, der entydigt identificerer entiteten i forhold til andre entiteter i samme entitetsæt. Er entiteten eksempelvis en patient, ville en oplagt primærnøgle være CPR-nummer, da dette ikke ændrer sig. Svage entiteter betegner derimod et entitetssæt, hvor attributterne ikke kan forme en primærnøgle. For at gøre et svagt entitetssæt meningsfuldt skal det være forbundet til et andet entitetssæt, kaldet det identificerende entitetssæt. Et svagt entitetssæt siges at være eksistensafhængig af det identificerende entitetssæt, og relationen mellem disse to kaldes derfor også for den identificerende relation. Nøgler bruges generelt til at specificere, hvordan entiteter i et givet entitetssæt kan skelnes fra hinanden. Hermed må to entiteter, i det samme entitetssæt, ikke have den præcist samme værdi for alle attributter. Kandidatnøgler udvælges nøje og kan anvendes som nøgler, men bliver det ofte ikke. Kandidatnøgler skal entydigt kunne identificere de enkelte entiteter, men må ikke kunne reduceres. (Silberschatz et al., 2001a)

Relationer er en association mellem flere entiteter som forbinder entiteterne/entitetssættene (fx. "Patient" *får* en "Diagnose"). Ligesom entiteter kan relationer også have attributter.

Figur 11.2 på næste side illustrerer ER-diagrammet for Biobankregisteret med entiteter,

primær- og kandidatnøgler, samt entiteternes relationer.



Figur 11.2: ER-diagram for Biobankregisteret med dertilhørende entiteter og deres relationer. Nøgleattributter er markeret med understregelse af teksten.

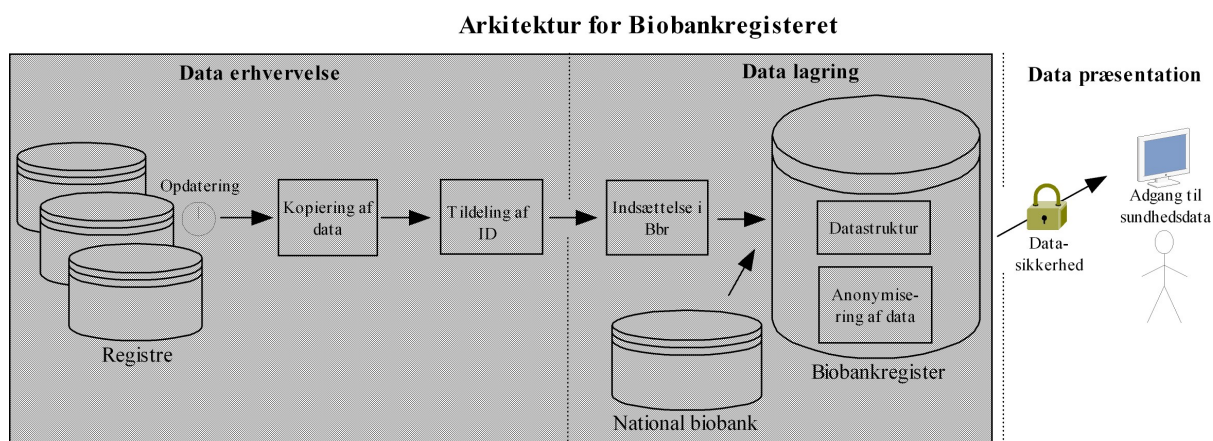
Biobankregisteret udgøres af ti entiteter: "Patient", "Kirurgisk behandling", "Medicinsk behandling", "Anden behandling", "Biologiske prøver", "Diagnose", "Mor/Far", "Børn", "Dødsårsag" og "Patologisvar". "Mor/Far" og "Børn" er begge svage entiteter, hvor den identificerende entitet er "Patient". Attributterne i ER-diagrammet stemmer ikke fuldstændig overens med tabelvariablerne for Landspatientregisteret (Sundhedsstyrelsen, 2010c), Patologiregisteret (Sundhedsstyrelsen, 2010d), Dødsårsagsregisteret (Sundhedsstyrelsen, 2010a) og Fødselsregisteret (Sundhedsstyrelsen, 2010b), men det kan de komme til ved en udvidelse af de specificerede attributter.

11.3 Anonymisering af relevante data

Anonymiseringen af data håndteres i ER-diagrammet ved at lagre det tildelte ID sammen med CPR-nummeret som en attribut til entiteten "Patient". Både CPR-nummer og ID er derfor også primærnøgler for entiteten.

Kapitel 12

Datapræsentation



Figur 12.1: Med udgangspunkt i figur 9.1 kan datapræsentationsprocessen for Biobankregisteret illustreres.

Som illustreret i figur 12.1 indebærer processen for datapræsentation hensyntagen til datasikkerhed og adgang til sundhedsdata. Der er flere forskellige metoder til at gøre information tilgængelig for forskere, der skal anvende Biobankregisteret. For forskeren er Biobankregisteret komponenten til levering af information, idet de via den online webportal kan få direkte indsigt i aggregerede sundhedsdata.

12.1 Datasikkerhed

Autenticitetssikring

Ved anvendelse af den online webportal, der giver forskere mulighed for at udtrække aggregerede data fra Biobankregisteret, skal der tages højde for datasikkerheden. Derfor er det relevant at anvende Den Gode Webservice (DGWS), der vha. standarder og arkitektur gør det muligt at udveksle data på tværs af systemplatforme via Internettet. Kommunika-

tion med udgangspunkt i DGWS involverer fire logiske operationer, som ligeledes gør sig gældende for kommunikationen mellem forsker og Biobankregisteret:

- Autentifikation: identificering af bruger
- Autorisation: godkendelse af bruger
- Forespørgsel: ønskede data forespørges
- Returnering: svar fra klientsystemet

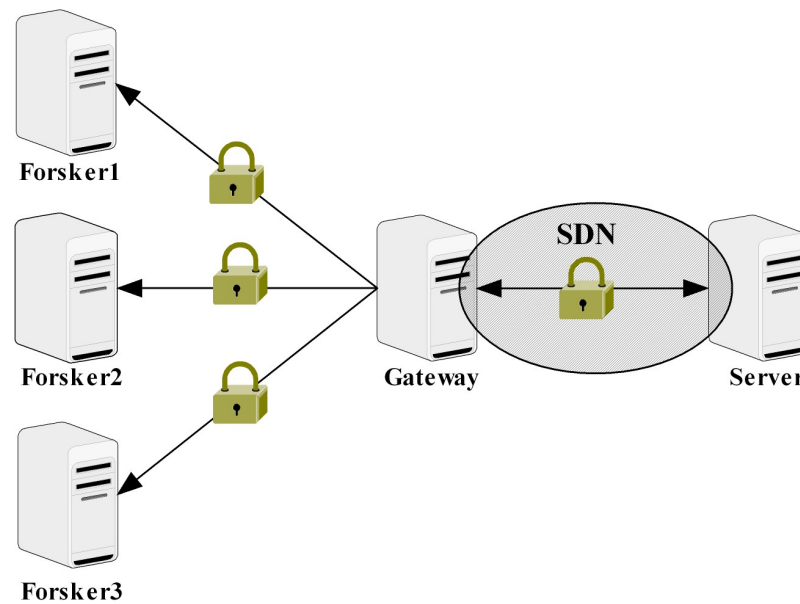
(MedCom og Sundhedsministeriet, 2008a)

Når der udveksles information mellem to parter, i dette tilfælde mellem Biobankregisteret og forskeren, vil informationens grad af følsomhed være afgørende for behovet for at bevise sin identitet. Meget følsomme data kræver meget pålidelige mekanismer, mens mindre følsomme data ikke kræver så stor sikkerhed. DGWS anvender IT- og Telestyrelsens vejledning omhandlende autenticitetssikring. Der er fire forskellige niveauer for tillid til brugerens identitet samt forskellige typer af akkreditiver der kan sikre de fire niveauer:

- Lille eller ingen autenticitetssikring → Ingen akkreditiver
- Nogen autenticitetssikring → Brugernavn og password
- Høj autenticitetssikring → Digital signatur på et "soft token" f.eks. installeret på en PC
- Meget høj autenticitetssikring → Digital signatur på et "hard token" f.eks. installeret på en USB nøgle

(MedCom et al., 2008)

En anden metode til kontrollering af systemets brugere er at anvende sikre netværk, der kun kan tilgås, ved at bruger og systemudbyder har indgået en kontrakt omkring anvendelsen af systemet. Et eksempel på sådan et netværk er MEDCOMs Sundhedsdatanet (SDN), hvortil mange sundhedsfaglige aktører er knyttet. Brugere af SDN skal underskrive en fysisk kontrakt omhandlende vilkårene for kommunikation mellem de to parter, for at kunne kommunikere via SDN. (MedCom og Sundhedsministeriet, 2008b) Nedenstående figur illustrerer SDN konfigurationen, hvor flere forskere anvender SDN.



Figur 12.2: SDN konfiguration hvor flere forskere anvender SDN. (MedCom og Sundhedsministeriet, 2008b)

Som beskrevet tidligere vil forskeren ikke kunne identificere patienten ud fra de patientoplysninger de modtager via den online webportal, og derfor kan graden af autenticitetssikring diskuteres. På baggrund af IT- og Telestyrelsens vejledning vedrørende niveauer af autenticitetssikring vurderes det, at en lille eller ingen autenticitetssikring er påkrævet. Derimod er det oplagt, at forskere kun kan opnå adgang gennem SDN, hvorfor der skal laves en kontrakt mellem forskere og MEDCOM, hvis en sådan ikke allerede eksisterer.

Sjældne sygdomme

Da nogle sygdomme kun rammer en meget lille del af befolkningen, kan det være nødvendigt at tage højde for dette i visualiseringen af aggregerede sundhedsdata fra den online webportal, da de enkelte patienter ellers kan være lette at identificere. Der skal derfor stilles et krav om, hvor mange patienter med en specifik diagnose, der som minimum skal eksistere for at data må visualiseres. Det samme gør sig gældende for behandlingstype.

EU definerer sjældne sygdomme som sygdomme, hvor prævalensen er 5 ud af 10.000 (Kommissionen for de europæiske fællesskaber, 2008). I dette projekt defineres en sjælden sygdom dog, som en sygdom der rammer ≤ 50 personer, da det vurderes, at patienter ikke kan identificeres hvis der er flere end 50 med en given sygdom. Det samme princip gør sig gældende for behandlingstype.

12.2 Adgang til relevante data

Brugeren kan få adgang til relevante sundhedsdata både fra den online webportal, hvor data er aggregerede, og ved at ansøge SSI om udtræk af personhenførbare sundhedsdata. Forskeren kan få adgang til den online webportal gennem et GUI, hvorfra der kan skabes et hurtigt overblik over aggregerede data samt de patienter, hvorpå der eksisterer biologiske prøver. Forskeren skal have mulighed for at udvælge følgende forskellige parametre: køn, aldersgruppe, diagnose, behandling, dødsårsag, familierelation, patologisvar og biologisk materiale.

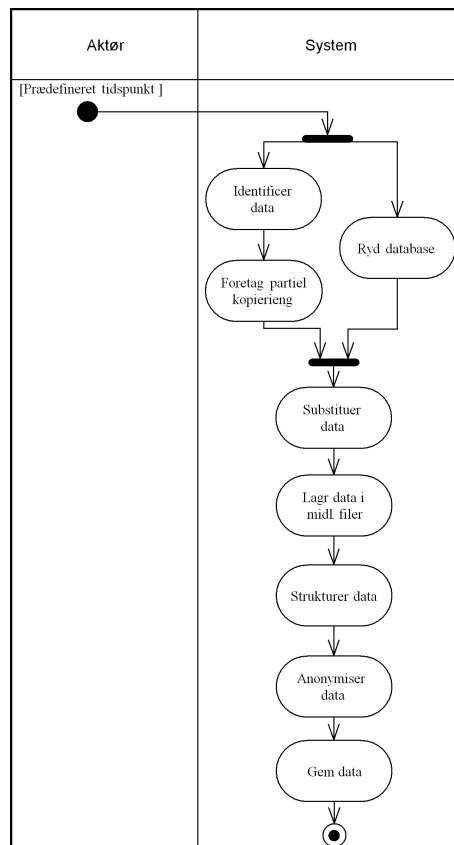
Anderledes forholder det sig, hvis forskeren ønsker udtræk af personhenførbare sundhedsdata. Her er det oplagt at SSI følger SST's fremgangsmåde, hvor forskeren skal lave en specificeret ansøgning til Datatilsynet, med en beskrivelse af forskningens formål. Efter modtaget accept fra Datatilsynet skal ansøgningen samt godkendelsen sendes til SSI, som efterfølgende laver udtræk fra Biobankregisteret. De udtrukne rå data vil blive lagret på en CD-rom, som efterfølgende vil blive sendt til forskeren med posten, og forskeren kan dermed anvende de forskellige rå data i forskellige programmer, se Appendiks C.

Kapitel 13

Aktivitetsdiagram med swimlanes

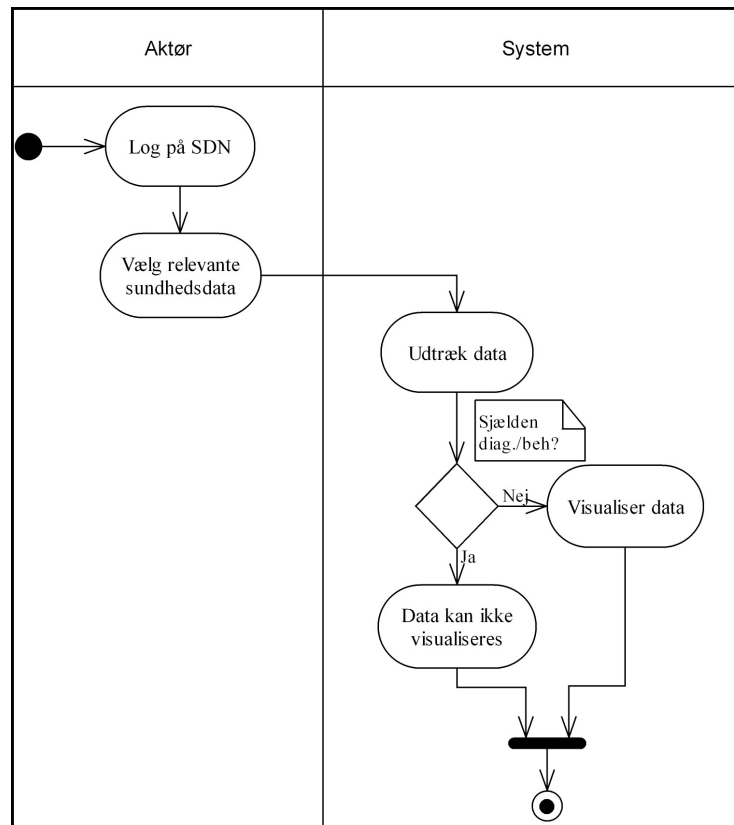
På baggrund af metoderne identificeret i Analysen kan aktivitetsdiagrammer med swimlanes illustreret i afsnit 8.3 udvides til følgende:

13.1 UC: "Kopier og gem data"



Figur 13.1: Aktivitetsdiagram med swim lanes for UC: "Kopier og gem data".

13.2 UC: "Visualiser data"



Figur 13.2: Aktivitetsdiagram med swim lanes for UC: "Visualiser data".

Del IV

DESIGN

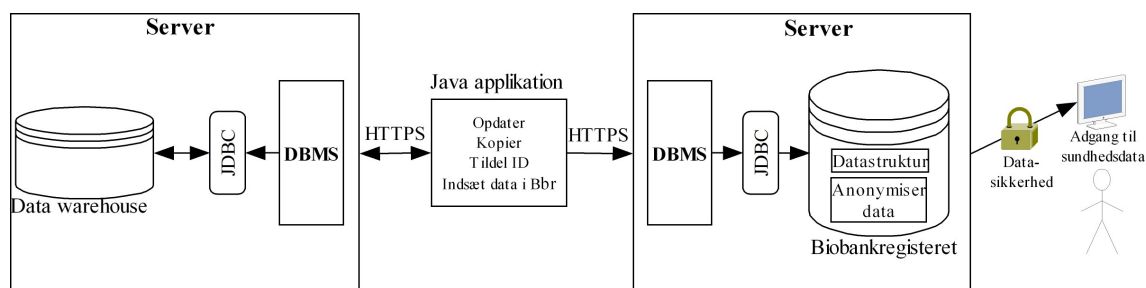
Design omhandler en specificering af standardmetoderne til dataerhvervelse, datalagring samt datapræsentation.

Kapitel 14

Systemets dele

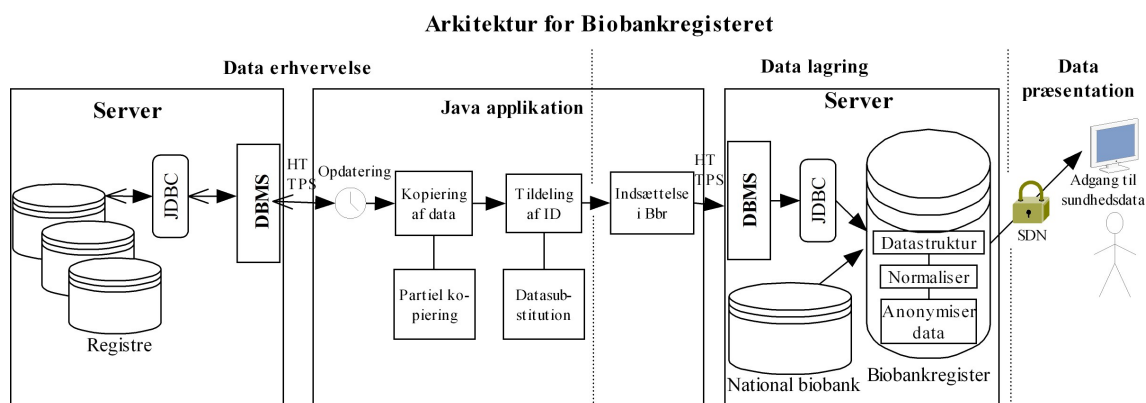
14.1 Systemarkitektur

Det blev i kapitel 8 i Krav beskrevet, at der skal designes en applikation, der håndterer opdatering, kopiering og tildeling af ID til de relevante sundhedsdata. Denne applikation skal derfor ligeledes håndtere kommunikationen mellem Biobankregisterets DBMS og det DBMS der styrer SST's registre. Applikationen kan programmeres i forskellige sprog som f.eks. Java, C og C++, hvoraf Java er det bedst egnede programmeringssprog, idet det er platformuafhængigt, og derfor kan fortolkes af alle computersystemer, der har en Javafortolker installeret. Derfor udbyder mange DBMS-leverandører en Java Database Connectivity (JDBC) driver, idet denne gør det muligt for Java applikationer at tilgå DBMS'et. For at Javaapplikationen skal kunne tilgå DBMS'et og JDBC driveren, skal der oprettes forbindelse til serveren og samtidig laves et JDBC funktionskald, hvorefter det er muligt for applikationen at tilgå databasen. (Elmasri og Navathe, 2003c) Selve kommunikationen foregår typisk over Internettet via Hypertext Transfer Protocol Secure (HTTPS), som er en protokol der ofte anvendes i forbindelse med tjenester, hvor kun afsender og modtager må kende meddelelsen (Beynon-Davies, 2000b). Med udgangspunkt i figur 8.1 fra Krav kan figur 14.1 på næste side, der illustrerer kommunikationen mellem systemets forskellige komponenter, udledes.



Figur 14.1: Overblik over kommunikationen mellem SST's datawarehouse og SSI's Biobankregister via applikationen, der håndterer opdatering, kopiering, tildeling af ID samt indsættelse af data i Biobankregisteret. Yderligere visualiseres forskerens adgang til Biobankregisteret. Fyldte pile illustrerer transport af relevante sundhedsdata, mens ufyldte pile illustrerer applikationens kommunikation med databasen.

Ud fra ovenstående figur kan figur 9.1 på side 54 fra Analysen uddybes til figur 14.2:



Figur 14.2: Systemarkitekturen for Biobankregisteret med de identificerede standardmetoder samt SST og SSI's servere, DBMS og JDBC drivere. Fyldte pile illustrerer transport af relevante sundhedsdata, mens ufyldte pile illustrerer applikationens kommunikation med databasen. Da opdatering og indsættelse af data i Biobankregisteret foretages af applikationen og ikke på baggrund af specifikke opdaterings- og indsættelsesmetoder, har ikke nogen specifikke metoder tilknyttet.

Der vil i dette projekt ikke fokuseres på et specifikt design af Biobankregisterets DBMS, da et DBMS ofte købes af større leverandører f.eks. IBM, Microsoft og Oracle. I stedet kan Appendiks F læses for yderligere indformation om standardkomponenterne, der er indeholdt i et DBMS.

Dataerhvervelse

Jævnfør figur 14.2 består dataerhvervelsen i at initiere opdateringen af Biobankregisteret, foretage en partiel kopiering af relevante sundhedsdata, og tildele CPR-nummeret et ID vha. datasubstitution, således at patienten får en ny nøgle hver gang der opdateres. Applikationen skal som beskrevet i ovenstående kapitel programmeres i Java, og da denne skal kommunikere med relationelle databaser, skal standardsproget SQL anvendes til at tilgå registrene samt Biobankregisteret.

15.1 Tidsinterval for opdatering

Det blev i Analysen valgt, at opdatering af databasen skal forekomme én gang ugentligt, hvilket håndteres af Java-applikationen. Dette kan gøres ved at lade applikationen kende til tid, det vil sige år, måneder, dage, timer og sekunder, og på den måde tillade opdatering med syv dages mellemrum. Når de relevante sundhedsdata skal lagres i Biobankregisteret, skal databaseskemaerne ryddes for data, hvilket kan håndteres med SQL-kommandoen `DELETE FROM`, der rydder alle rækker i en tabel uden at slette tabelstruktur, -attributter og -indeks (Elmasri og Navathe, 2003d).

15.2 Kopiering af relevante data

I Analysen blev det klarlagt, at en partiel kopiering af de relevante sundhedsdata kan håndteres af en Java-applikation. Da SST anvender et relationelt DBMS vil kopieringen af data foregå ved brug af SQL-kommandoen `SELECT`. Denne udvælger de relevante sundhedsdata i de specifikke tabeller i registrene, således at de udelukkende kopieres og dermed stadig figurerer i SST's registre, samtidig med at de bliver lagret i Biobankregisteret. (Elmasri og Navathe, 2003e) Herefter skal de udtrukne relevante sundhedsdata lagres i en midlertidig fil med henblik på lagring i Biobankregisteret. Filen skal derfor transporteres

fra SST's data warehouse til SSI's Biobakregister via Internettet, hvorfor der skal oprettes kommunikation imellem SST's server og applikationen samt applikationen og SSI's server.

15.3 Tildeling af ID

Det blev i Analysen valgt at anvende datasubstitution, hvorfor applikationen automatisk skal håndtere tildelingen af et ID til alle erhvervede CPR-numre, hver gang databasen opdateres, således at hver patient er unikt identificeret i databasen. Antallet af ID der skal tildeles kan findes vha. en for løkke der tæller antallet af CPR-numre der udtrækkes.

Kapitel 16

Datalagring

Dette kapitel omhandler indsættelse af de kopierede relevante sundhedsdata i Biobankregisteret, normalisering af databaseskemaer og anonymisering af data i databasen, se figur 14.2 på side 72. For at lagre de relevante sundhedsdata omsættes ER-diagrammet fra Analysen til en samling af relationer, der udgør designstrukturen for Biobankregisteret. Ud fra ER-diagrammet konstrueres tabeller, som efterfølgende skal normaliseres for at danne relationerne. I denne proces kan anonymiseringen af de relevante sundhedsdata håndteres.

16.1 Indsættelse af relevante data i Biobankregisteret

For at indsætte de relevante sundhedsdata i Biobankregisteret skal applikationen anvende SQL-kommandoen `INSERT INTO`, hvormed attributterne indsættes i de rigtige tabeller. (Elmasri og Navathe, 2003f)

16.2 Datastruktur

Det overordnede design af databasen kaldes databaseskema. Konceptet med databaseskemaer kan forstås ved en analogi til et program skrevet i et programmeringssprog. Databaseskemaet svarer til variabeldeklareringen i et program, og variabelværdierne i programmet ved et givent tidspunkt svarer til en instans af et databaseskema. (Silberschatz et al., 2001b)

Databaseskemaer

Databaseskemaet udledes fra ER-diagrammet med det formål at reducere ER-diagrammet til en samling af relationer, der udtrækker det samme som ER-diagrammet. Ved at konstruere skemaerne kan information lagres uden unødvendig redundans, samtidig med at information let kan genfindes, hvilket hjælper til effektiv forespørgselsprocessering. (Sil-

berschatz et al., 2001c) For at opnå dette anvendes normalisering, da det hindrer uhensigtsmæssigheder i de resulterende relationer. Normalisering foregår i tre forskellige trin: Første, anden og tredje normalform. I første normalform må der ikke forekomme gentagelser af elementer eller grupper af elementer, dvs. attributterne må ikke være delelige. En relation R er derfor på første normalform, når domænet for alle attributter for R er atomare. Anden normalform opnås, hvis en relation R er på første normalform, og hver ikke-nøgleattribut er fuldt funktionelt afhængig af hver kandidatnøgle i R . Tredje normalform opnås, hvis en relation R er på anden normalform og ikke har nogen transitiv afhængighed af ikke-nøgleattributter. (Silberschatz et al., 2001d)

Relationerne visualiseret nedenfor er på hhv. første, anden og tredje normalform. Understregede ord visualiserer primærnøglerne for den specifikke relation.

Patient

Patient	Patientid	Patient
<u>id</u>	<u>patient.id</u>	<u>id</u>
cprnr	cprnr	fødselstilstand
fødselstilstand		
(a)	(b)	(c)

Tabel 16.1: (a) Første og anden normalform for entiteten Patient, (b) tredje normalform for $\text{patient.id} \rightarrow \text{cprnr}$ (c) tredje normalform for entiteten Patient.

Kirurgisk, Medicinsk og Anden behandling

Kbehandling	Mbehandling	Abehandling
<u>specifikation</u>	<u>medicintype</u>	<u>specifikation</u>
varighed	<u>dosis</u>	varighed
	varighed	
(a)	(b)	(c)

Tabel 16.2: Første, anden og tredje normalform for entiteten (a) Kirurgisk behandling, (b) Medicinsk behandling og (c) Anden behandling.

Biologiske prøver, Dødsårsag og diagnose

Biologiskeprøver	Dødsårsag	Diagnose
<u>specifikation</u>	<u>specifikation</u>	<u>specifikation</u>
type	<u>navn</u>	(c)
(a)	(b)	

Tabel 16.3: Første, anden og tredje normalform for (a) entiteten Biologiske prøver, (b) entiteten Dødsårsag og (c) entiteten Diagnose.

Patient_Kbehandling, Patient_Mbehandling og Patient_Abehandling

Patient_Kbehandling	Patient_Mbehandling	Patient_Abehandling
<u>kbehandling.specifikation</u>	<u>mbehandling.dosis</u>	<u>abehandling.specifikation</u>
<u>patient.id</u>	<u>mbehandling.medicintype</u>	<u>patient.id</u>
dag	<u>patient.id</u>	dag
måned	dag	måned
år	måned	år
(a)	(b)	(c)

Tabel 16.4: Første, anden og tredje normalform for relationen (a) mellem entiteterne Kirurgisk behandling og Patient, (b) mellem entiteterne Medicinsk behandling og Patient og (c) mellem entiteterne Anden behandling og Patient.

Patient_Bprøver, Patient_Dødsårsag og Patient_Diagnose

Patient_Bprøver	Patient_Dødsårsag	Patient_Diagnose
<u>bprøver.specifikation</u>	<u>dødsårsag.specifikation</u>	<u>diagnose.specifikation</u>
<u>patient.id</u>	<u>patient.id</u>	<u>patient.id</u>
dag	dag	dag
måned	måned	måned
år	år	år
(a)	(b)	(c)

Tabel 16.5: Første, anden og tredje normalform for relationen (a) mellem entiteterne Biologiske prøver og Patient, (b) mellem entiteterne Dødsårsag og Patient og (c) mellem entiteterne Diagnose og Patient.

Børn og Patologisvar

Børn	Børn	Patologisvar
<u>patient.id</u>	<u>patient.id</u>	<u>materialebeskrivelse</u>
<u>id</u>	<u>id</u>	diagnose
<u>cprnr</u>		kommentarer
(a)	(b)	(c)

Tabel 16.6: (a) Første og anden normalform for entiteten Børn, (b) tredje normalform for entiteten Børn, og (c) første, anden og tredje normalform for entiteten Patologisvar.

Mor/Far og Patient_Patologisvar

Mor/Far	Mor/Far	Patient_Patologisvar
<u>patient.id</u>	<u>patient.id</u>	<u>patologisvar.mbeskrivelse</u>
<u>id</u>	<u>id</u>	<u>patient.id</u>
<u>cprnr</u>		dag
(a)	(b)	måned
		år
		(c)

Tabel 16.7: (a) Første og anden normalform for entiteten Mor/Far, (b) tredje normalform for entiteten Mor/Far og (c) første, anden og tredje normalform for relationen mellem entiteterne Patient og Patologisvar.

16.3 Anonymisering af relevante data

Ved ovenstående normalisering af relationerne foretages den ønskede anonymisering implicit, idet CPR-nummer og ID samles i egen tabel for at bringe entiteten Patient på tredje normalform. Dermed bliver kun ID nøgle for entiteten Patient, og det bliver muligt at koble patientdata med behandling, diagnose osv. via det tildelte ID. Nedenstående tabel visualiserer entiteten Patientid, der ligger til grund for anonymiseringen.

Patientid
<u>patient.id</u>
cprnr

Tabel 16.8: Tabellen, hvormed anonymisering udføres.

Ønskes der viden om, hvilket ID der tilhører et bestemt CPR-nummer kan SQL-kommandoen NATURAL JOIN anvendes.

Datapræsentation

Datapræsentationen består i at forskere skal logge på SDN for at få visualisere relevante sundhedsdata i et GUI, mens forskere ikke skal have mulighed for at få indblik i sjældne sygdomme, jævnfør figur 14.2 på side 72.

17.1 Datasikkerhed

Autenticitetssikring

I Analysen blev det besluttet, at forskere skal have adgang til SDN for at anvende den online webportal. Forskere kan få adgang til SDN ved brug af egen specifik IP-adresse. For at overholde sikkerhedsforskrifterne for SDN skal den webserviceklient og -server, som adressen benyttes af, forpligtige sig til at beskytte eget netværk. Derfor skal forskere, der benytter SDN til webservicekommunikation overholde sikkerhedsforeskrifterne. (MedCom og Sundhedsministeriet, 2008b)

Da forbindelsen til SDN vil blive brugt af flere forskere og dermed flere systemer, er det ikke længere så gennemskueligt, hvorvidt konfidentialiteten er sikret fra klienten til serveren, samt hvilke systemer der er opkoblet til SDN. Derfor bør der anvendes en VPN-klient, der etablerer forbindelse mellem den server, der håndterer adgangen til SDN, og forskerens decentrale it-system. Vælges det at følge DGWS skal forbindelsen mellem systemet, der varetager adgangen til SDN, og forskerens decentrale it-system etableres via VPN eller SSL/TLS, som er en protokol, der sikrer konfidentialitet mellem to systemer, der kommunikerer via HTTP over internettet. Både VPN og SSL/TLS kan derfor anvendes i dette tilfælde. (MedCom og Sundhedsministeriet, 2008b)

Sjældne sygdomme

Det blev i Analysen besluttet at sjældne sygdomme defineres, som en sygdom der rammer ≤ 50 personer. Samme princip gør sig gældende for behandlingstype. Der skal derfor være en funktion i Java applikationen, der kan tælle antallet af patienter med samme diagnose eller behandlingstype efterhånden som de udtrækkes af databasen, hvilket kan klares med en for løkke. For at afgøre om data kan visualiseres, kan der anvendes en if-else sætning, der returnerer 0 hvis antallet er under 50 og 1 hvis antallet er over 50. Returneres tallet 0 informeres brugeren om, at sygdommen er for sjælden, og at data ikke kan visualiseres.

17.2 Adgang til relevante data

I Analysen i afsnit 12.2 på side 66 blev det beskrevet, hvilke anonymiserede data der skal kunne udtrækkes og få indsigt i via den online webportal. Ligeledes blev det valgt at adgang til sundhedsdata kun kan opnåes via SDN. Forskeren vil herfra få mulighed for at udvælge forskellige parametre, hvilket betyder at der kan foretages en bred kombination af søgninger, der ikke på forhånd er kendte. Derfor skal der anvendes et ad hoc rapporteringssystem. Ad hoc rapporteringsprocessen består af tre trin:

- Forespørgsel: udtræk af relevante data
- Rapportering: formatering af data så det kan præsenteres hensigtsmæssigt
- Analyse: analyse der foretages på dataudtrækket.

Skærmbillederne kan udarbejdes som en JAVA webapplikation (JApplet), da denne kan afvikles på ethvert styresystem. (Elmasri og Navathe, 2003c) For at konstruere denne webapplikation skal der bl.a. anvendes JButton, JComboBox og JTextField for at visualisere knapper, scrollbar, tekstfelter osv. I det følgende gives et mock-up eksempel på, hvorledes skærmbillederne i den online webportal kan designes.

Velkomst GUI

Det første forskeren introduceres for ved anvendelse af den online webportal, er et skærm-billede, hvorfra det er muligt at vælge, hvilke data der ønskes viden om. Da forskeren kan ønske at studere kønsrelaterede sygdomme, skal det være muligt at vælge køn (mænd, kvinder, begge) ved hjælp af en "drop-down" knap. Det er ligeledes hensigtsmæssigt, at forskeren ved brug af en "scrollbar" kan vælge at få indblik i data i alle eller specifikke aldersgrupper, idet forskeren eksempelvis kan have én bestemt aldersgruppe som sit fokus. To centrale parametre er diagnose og behandling. Forskeren kan have ønske om at få information om patienter med en specifik diagnose, som derfor skal indtastes vha. diagnosekoder. Når

denne er indtastet, skal de behandlinger, som diagnosen tilhører, automatisk visualiseres ved hjælp af en "drop-down" knap, og forskeren skal have mulighed for at vælge én, to eller alle behandlinger på netop denne diagnose. Det kan ligeledes være fordelagtigt for forskeren at få indblik i om der eksisterer patologisvar og biologiske prøver på patienterne indenfor den specificerede aldersgruppe med en specifik diagnose, hvorfor forskeren skal kunne vælge disse. Endvidere er behandlingsoutcome en central parameter, idet den fortæller, hvorvidt patienterne er døde eller levende, og forskeren skal derfor kunne udvælge denne. Forskeren skal ligeledes kunne få visualiseret, hvor mange familier der er ramt af en specifik diagnose, hvorfor familierelation også skal kunne vælges. Efter valg af parametre trykkes "OK" og forskeren får visualiseret data i aggregeret tilstand.

I nedenstående figur 17.1 ses et eksempel på, hvordan det første skærm billede i den online webportal kan se ud.

Figur 17.1: Mock-up af velkomstskræmbilledet, hvorfra forskeren har mulighed for at udvælge forskellige sundhedsdata.

I det tilfælde at brugeren ønsker at få visualiseret aggregerede sundhedsdata for en sjælden sygdom, vil der fremkomme en meddelelse til brugeren om at data ikke kan visualiseres og velkomst GUI forbliver visualiseret.

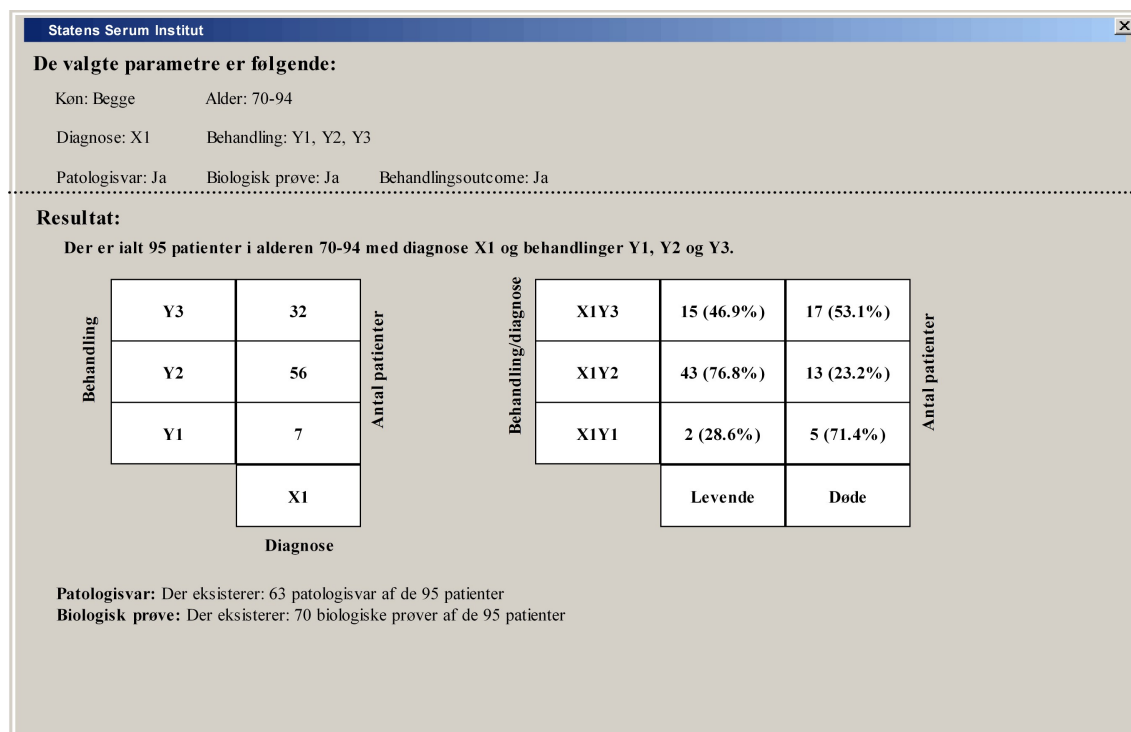
Aggregerede data GUI

Når forskeren har udvalgt sine parametre og denne har trykket "OK" skal et nyt skærm-billede med de aggregerede parametre visualiseres. Øverst i billedet skal forskerens valg

illustreres, så disse er synlige samtidig med visning af resultaterne. Derudover skal resultaterne illustreres for forskeren ved anvendelse af tabeller. I figur 17.2 ses et eksempel på dette skærbillede med visning af forskerens valg samt resultaterne deraf. Forskeren har valgt følgende:

- Køn: begge
- Alder: 70-94
- Diagnose: X1
- Behandling: Y1, Y2, Y3
- Patologisvar: Ja
- Biologiske prøver: Ja
- Behandlingsoutcome: Ja

Ved valg af disse parametre vil der blive oprettet forbindelse til Biobankregisteret, som indeholder data om alle disse parametre, og data vil blive udtrukket og visualiseret for forskeren som aggregerede sundhedsdata. Derudover vil antallet af patologisvar samt biologiske prøver blive skrevet på skærbilledet.



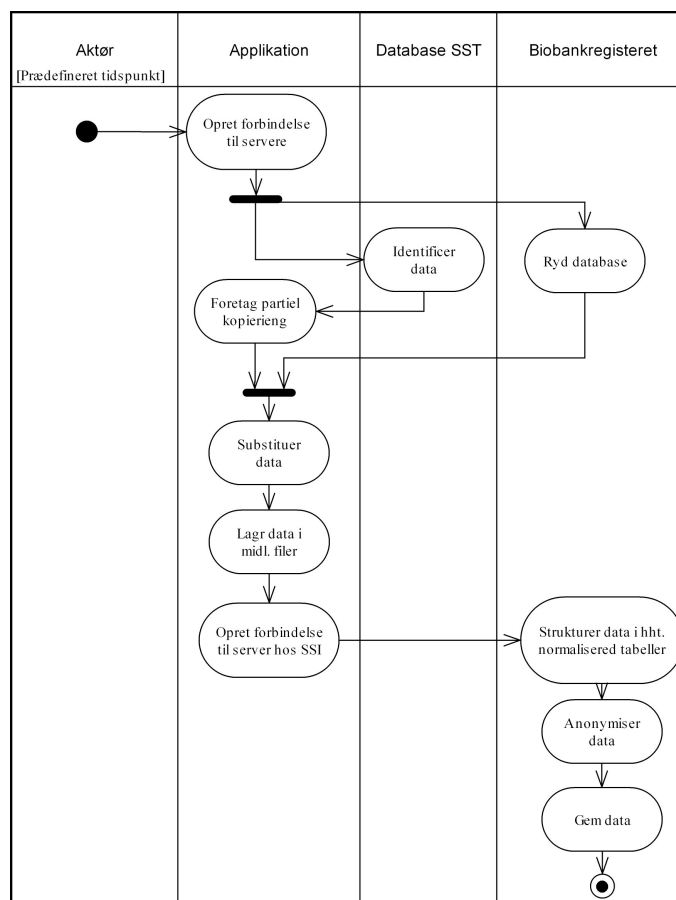
Figur 17.2: Mock-up af skærbilledet der visualiserer de udvalgte sundhedsdata som aggregerede data for forskeren.

Kapitel 18

Aktivitetsdiagram med swimlanes

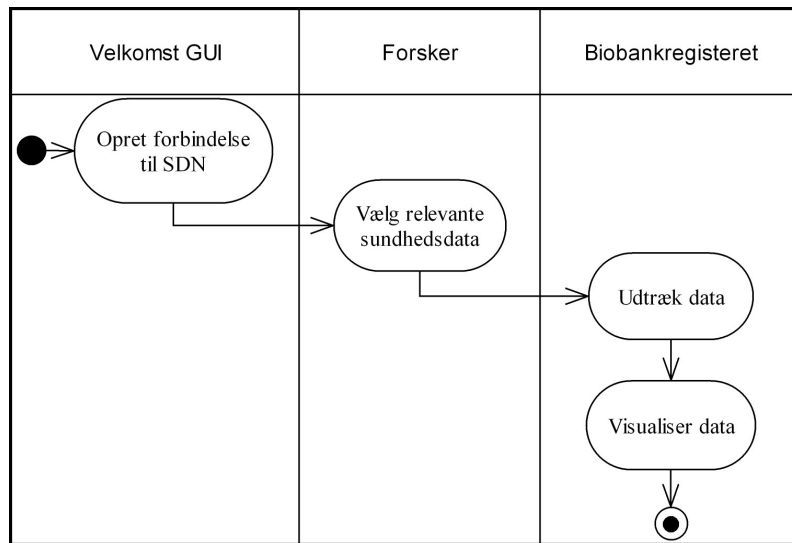
På baggrund af metoderne identificeret i Design kan aktivitetsdiagrammer med swimlanes illustreret i afsnit 8.3 udvides.

18.1 UC: "Kopier og gem data"



Figur 18.1: Aktivitetsdiagram med swim lanes for UC: "Kopier og gem data".

18.2 UC: "Visualiser data"



Figur 18.2: Aktivitetsdiagram med swim lanes for UC: "Visualiser data".

Del V

EVALUERING og KONSEKVENSVURDERING

I denne del vil det valgte kompromis, systemarkitekturen og systemets tekniske indhold blive evalueret. Yderligere vil der foretages en konsekvensvurdering.

Evaluering

Løsningsstrategien for dette projekt var at skitsere en struktur, der understøtter et kompromis for datahåndteringen i den nationale biobank. Yderligere ønskedes det at skitsere forskerens interaktion med systemet. Visionen var at foretage datahåndteringen vha. tekniske standardmetoder. Derfor evalueres kompromiset, systemarkitekturen, samt det tekniske indhold. På grund af tidsrammen for projektet evalueres brugerens interaktion med systemet ikke. Da systemet udelukkende er designet og dermed ikke implementeret, er det ikke muligt at foretage en systemtest. I stedet kan selve konceptet med systemet evalueres ved en koncepttest. Koncepttesten deles op i tre: kompromis, systemarkitektur og teknisk indhold.

19.1 Metode til koncepttest

Evalueringen af kompromis, systemarkitektur og teknisk indhold foretages ved at vise materiale til relevante interessanter og stille dem spørgsmål ang. de tre evalueringsområder som de bedes besvare, hvorfor evalueringen er af kvalitativ karakter. Dette er meget lig proceduren ved en koncepttest, der udføres ved at visualisere ideer for systemet samt systemkonceptet for interessanter og brugere af systemet. Herefter stilles relevante spørgsmål, hvortil de bedes tilkendegive deres vurdering eller mening. Derfor udføres denne evaluering som en koncepttest. Formålet med denne koncepttest er at få en vurdering af både kompromis, systemarkitektur og det tekniske indhold/standardmetoderne. Dette er ønskværdigt i projektet, da det giver en fornemmelse for, hvorvidt Løsningstrategien og Visionen er velegnet for datahåndteringen i den nationale biobank. Valget af de tre evalueringsområder sker med baggrund i Løsningsstrategien som var at skitsere arkitekturen for et system, der understøtter hele datahåndteringsprocessen og dermed det valgte kompromis.

En koncepttest kan eksekveres ved forskellige metoder såsom et personligt interview, telefonisk interview og via emailkorrespondence. Personligt interview giver mulighed for at

se brugerens reaktioner på det visualiserede koncept samt føre en samtale med personen, hvor spørgsmål, der ønskes besvaret, kan stilles. Telefonisk interview er anvendeligt, hvis det ønskes at føre en samtale med respondenterne, og spørgsmålene kan besvares uden behov for visualisering af konceptet. Emailkorrespondence giver mulighed for visualisering, men stiller større krav til de stillede spørgsmål, idet der ikke er mulighed for at stille uddybende og forklarende spørgsmål løbende. På grund af projektets tidsfrist har det ikke været muligt at foretage personlige interviews. Det vurderes, at et telefonisk interview ikke vil give respondenterne den ønskede mulighed for at få konceptet visualiseret, hvorfor emailkorrespondence er den bedste metode. Derfor vælges emailkorrespondence, til trods for at det stiller større krav til de stillede spørgsmål, da det giver respondenterne mulighed for at få visualiseret systemets arkitektur og anvendte metoder, således de kan foretage en grundig vurdering af kompromiset, systemarkitekturen og det tekniske indhold.

Kompromis

For at evaluere kompromiset for datahåndteringen, der omhandler adgang til registerdata og biologiske prøver fra én platform, ønskes det at få en vurdering af kompromisets funktionalitet fra en medarbejder fra SST's dokumentationsenhed. Dette skyldes at SST skal være leverandør af sundhedsdata, og at konceptet ligeledes medfører flest konsekvenser for dem. Evalueringen er udført ved udveksling af spørgsmål og svar via email, hvor de tre kompromiser for datahåndteringen præsenteres for respondenterne i en tilsvarende form som i dette projekt, se afsnit 5. Herefter bedes respondenterne vurdere, hvilket kompromis der er det mest funktionelle, om det valgte kompromis kan udføres og herunder om det kan udføres vha. kopiering. Spørgsmål og svar kan ses i Appendiks G.

Systemarkitektur

Ligeledes ønskes det at få en vurdering af, hvorledes systemarkitekturen og indholdet kan være en løsning for den nationale biobank. Derfor bedes ansvarshavende for projektet hos SSI vurdere systemarkitekturen, illustreret i figur 9.1 på side 54, idet de har opstillet visionerne for den nationale biobank, se afsnit 4.1, og derfor som udgangspunkt må have gjort sig nogle tanker om systemarkitekturen. Ligesom i det ovenstående er evaluering udført ved udveksling af spørgsmål og svar over email. Respondenterne bedes svare på spørgsmål, ud fra illustrationen, der adresserer deres vurdering af den valgte systemarkitektur som en løsning for den nationale biobank, samt hvordan en alternativ systemarkitektur kunne se ud. Se Appendiks G.

Teknisk indhold

For at teste det tekniske indhold i systemarkitekturen blev en chef arkitekt fra CSC, der leverer systemløsninger i samme størrelsesorden, som dette koncept ligger op til, bedt om at vurdere det tekniske indhold og dermed de anvendte standardmetoder, med henblik på om standardmetoderne er teknisk gangbare. Ligesom i de forrige afsnit er evalueringen udført ved udveksling af spørgsmål og svar via email, hvor standardmetoderne præsenteres for respondenten ved at sende et uddrag af projektet, omhandlende Analyse og Design af systemkonceptet, inklusiv visualiseringer af dataarkitekturen. Heri bedes respondenten om at besvare spørgsmål, der adresserer den tekniske gangbarhed af disse standardmetoder samt hvilke metoder der kunne være anvendt i stedet.

19.2 Resultat

Kompromis

En samtale med en medarbejder fra SST's dokumentationsenhed, foretaget på baggrund af den sendte email, bekræftede at det valgte kompromis var det bedste af de tre, idet der sker en arbejdsdeling således ikke kun SST eller SSI står med ansvaret for både sundhedsdata og den nationale biobank. Ligeledes bevarer SST deres funktion som registeransvarlig, samtidig med at SSI kan få adgang til de relevante sundhedsdata. SST mener dog ikke, at kopiering af data er en god løsning, idet de har specialistviden om dataindhold og -definitioner, hvilket SSI ikke har, hvorfor de ikke kan håndtere de data.

Ud fra samtalen blev det ligeledes udledt, at SST ikke er enig i den opstillede systemarkitektur, idet de ikke mener, at kopiering er en anvendelig metode til erhvervelse af data hos SSI. SST lægger dog vægt på vigtigheden i, at det skal være muligt for forskere at få let adgang til både sundhedsdata og biologiske prøver, og at dette burde være muligt på den ene eller anden måde. Se Appendiks G

Systemarkitektur

Emailkorrespondence med Mads Melbye bekræftede, at den opstillede systemarkitektur er meget lig hvad der er tænkt for den nationale biobank: "*Ideen er at indhente få nøgleinfo fra de respektive biobanker til det nationale Biobankregister. Registeret trækker endvidere fra sundhedsstyrelsen informationer fra sundhedsregistre etc. således at du har den "komplette" information på de enkelte prøver. I aggregeret form lægges disse data på internettet således at forskere kan søge på hvor mange og hvilke prøver der f.eks. eksisterer på sklerose patienter. Ved henvendelse til det koordinerende center kan man så få rådgivning om hvordan man søger (vid.etisk kommite, etc.) for at få den reelle adgang til det biologiske materiale.*". Se Appendiks G

Derfra kan det uddrages, at den valgte systemarkitektur kan være en løsning for den nationale biobank, og at der ikke umiddelbart er en anden systemarkitektur, ifølge SSI, der kunne anvendes i stedet.

Yderligere kan det uddrages, at SSI også indirekte støtter det valgte kompromis, identificeret i Løsningsstrategien, da systemarkitekturen understøtter dette kompromis. Ligeledes indvendes der ikke mod de anvendte standardmetoder, som visualiseres i systemarkitekturfiguren, hvorfor det kan tænkes at disse metoder er i tråd med SSI's tanker for datahåndteringen.

Teknisk indhold

Samtale med chef arkitekt John Møller-Jensen bekræftede at de identificerede standardmetoder kan anvendes til udførelse af systemkonceptet, men at de er meget pragmatiske. Opdatering af Biobankregisteret, hvor alle tabeller ryddes og data indsættes på ny, kan foretages på en mere effektiv måde ved kun at kopiere de nyeste data. Dette er dog også en dyrere løsning. Tildelingen af et nyt ID ved hver opdatering kan være uhensigtsmæssig for forskeren, idet det vil være umuligt at koble gamle og nye dataudtræk. Se Appendiks G

Yderligere mente John Møller-Jensen, at der bør være forbindelse til CPR-registeret, idet dette via CPR-nummeret folkeregistrerer enhver på deres bopælsadresse (Det Centrale Personregister, 2010). Hermed kan patientens CPR-nummer altid identificeres, hvis der er viden om patientens bopæl, og der er dermed ikke afhængighed af at sygehusene indberetter et korrekt CPR-nummer.

19.3 Forbehold

Ovenstående metode til udførelse af koncepttest er sammensat af projektgruppen selv og derfor ikke en anerkendt metode. Skulle koncepttesten udføres på bedste vis burde brugerne af systemet, i dette tilfælde forskerne, også kontaktes, da de kan give en vurdering af systemets funktionalitet og brugervenlighed. Yderligere burde koncepttesten udføres løbende for at revurdere konceptet inden det endelige design. Samtidig er der kun kontaktet én respondent fra hver interessent til at validere systemkonceptet, hvorfor evalueringen er subjektiv.

19.4 Konklusion

Koncepttesten viste, at det valgte kompromis er i god tråd med SSI's forestilling om hvordan arbejdsfordelingen for håndtering af data i den nationale biobank skal håndteres. Derimod mener SST ikke, at kopiering af data og dermed systemarkitekturen er en løsning

idet SSI ikke har den nødvendige viden om dataindhold og -definitioner.

Ligeledes viste koncepttesten at systemarkitekturen er i god tråd med SSI's tanker og visioner for den nationale biobank, og at det kunne være en løsning for implementeringen af den nationale biobank. Dermed kan det også udledes, at det valgte kompromis er i god tråd med SSI's visioner, idet systemarkitekturen understøtter kompromiset.

De identificerede standardmetoder er teknisk gangbare, omend de er meget pragmatiske, hvorfor de opstillede use cases er opfyldt. Det blev også vurderet, at opdateringen af Biobankregisteret kan udføres, så kun de nyeste sundhedsdata blev lagret, hvorved CPR-nummeret ligeledes ikke tildeles et nyt ID hver uge, hvilket er fordelagtigt for forskerne i forhold til at kunne koble nye og gamle udtræk.

Konsekvensvurdering

I Løsningsstrategien blev der identificeret et kompromis, der tillader kobling af relevante sundhedsdata og biologiske prøver på en fælles platform, samtidig med at det giver færrest konsekvenser for SSI og SST. Derfor består designet af systemkonceptet som beskrevet i at kopiere, anonymisere, og lagre relevante sundhedsdata fra SST's data warehouse i Biobankregisteret, hvilket er fordelagtigt, idet forskere i samme arbejdsgang kan få adgang til sundhedsdata samt biologiske prøver. Dette kapitel fokuserer derfor på de konsekvenser, der vil opstå efter implementering af den nationale biobank samt det udviklede systemkoncept for både SST og SSI. Ligeledes vil der være en vurdering af, hvorvidt det identificerede kompromis er realistisk at udføre. Relevante konsekvenser indbefatter arbejdsdelingen imellem SST og SSI samt de økonomiske forhold, hvorfor der vil være fokus på disse i denne konsekvensvurdering. Arbejdsdelingen er relevant, idet den nationale biobank medfører centrale ændringer i begge parter arbejdsprocesser i forbindelse med datahåndteringen efter implementering af den nationale biobank. Økonomiske forhold er relevante, da driften af den nationale biobank vil have omkostninger for både SSI og SST. Ud fra en identificering af konsekvenserne vedrørende arbejdsdeling og økonomi kan relevante forhold, der afgør, hvorvidt det er hensigtsmæssigt at indføre kompromiset, eller om der er nogle u hensigtsmæssige konsekvenser, udledes.

20.1 Arbejdsdeling for datahåndtering

Implementering af dette systemkoncept for Datahåndteringssystemet vil have en effekt på arbejdsdelingen mellem SST og SSI, idet de skal beslutte, hvorledes arbejdet med sundhedsdata skal fordeles imellem dem. Dertil hører en stillingtagen til, samt enighed om, hvorledes sagsgangen ved dataudtræk skal foregå. I den forbindelse kan to scenarier for dataudtræk identificeres:

1. Forskere kan ansøge både SST og SSI om dataudtræk.

2. Opsættelse af restriktioner kan specificere, at forskere skal henvende sig til SST for dataudtræk, mens SSI kan ansøges, hvis der ønskes dataudtræk og information om biologiske prøver samtidig.

SST

Ovenstående scenarier vil begge medføre nogle konsekvenser for SST. Hvis det er muligt at ansøge om dataudtræk hos såvel SST som SSI, er der risiko for at mange forskere vil ansøge SSI om dataudtræk, og dermed mister SST's Forskerservice mange af sine funktioner, mens SSI vil overtage en del af det arbejde Forskerservice udfører. En anden konsekvens der kan have indlydelse på antallet af ansøgninger hos SST er, hvis det implementerede system hos SSI kan yde en bedre service end Forskerservice hos SST. Er dette tilfældet vil flere forskere eventuelt ansøge SSI om dataudtræk, netop fordi de har muligheden for det.

Ved oprettelse af restriktioner for dataudtræk vil SST's Forskerservice miste et fåtal af sine funktioner, idet forskere udelukkende vil henvende sig til SSI, når de ønsker indsigt i biologiske prøver og sundhedsdata på samme tid. Dermed vil forskere stadig kontakte SST, hvis de kun ønsker indsigt i data, og der vil derfor kun være et begrænset fald i antallet af forespørgsler hos Forskerservice. Ydermere vil SST fortsat være registeransvarlig, idet Biobankregisteret vil modtage kopierede sundhedsdata, og dermed skal SST fortsat varetage kontakten til sygehusene, og i samspil med dem sikre, at indberetningen af data til SST er optimal. Derfor er SST's funktion stadig vigtig for de forskere, der ønsker indsigt i andre data end de relevante sundhedsdata.

En yderligere konsekvens for SST vil være, at de skal hjælpe SSI med at vedligeholde viden om de relevante sundhedsdata, som befinder sig i Biobankregisteret, hvilket kan medføre en organisatorisk ændring samt en ekstra arbejdsbyrde for SST.

SSI

Hvis det første scenarie bliver en realitet efter implementeringen af den nationale biobank, vil det ligeledes medføre nogle konsekvenser for SSI, omend ikke i samme omfang som SST. Dette skyldes, at forskere har mulighed for at ansøge SSI om dataudtræk, hvilket muligvis ikke var SSI's hensigt med den nationale biobank, og de vil derfor få mere arbejde med dataudtræk end forventet og ønsket, samtidig med at de skal varetage flere opgaver. Omvendt vil oprettelse af restriktioner være en fordel for SSI, idet de kan yde den hjælp for forskerne, som de ønsker ved at levere information om både relevante sundhedsdata, stamdata om biologiske prøver samt de faktiske biologiske prøver.

20.2 Økonomi

De økonomiske konsekvenser ved implementering af den nationale biobank skal SST og SSI ligeledes tage højde for.

SST

Implementering af den nationale biobank vil have den konsekvens, at SST's Forskerservice vil miste nogle af deres funktioner og dermed vil Forskerservice' indkomst fra dataudtræk reduceres, idet de aflønnes pr. time, der skal bruges på at udtrække data (Sundhedsstyrelsen, 2010e).

SSI

Når den nationale biobank er implementeret vil SSI have nogle konstante driftomkostninger forbundet ved dette heraf både tekniske og personalemæssige omkostninger. Ydermere skal SSI aflønne personale fra SST, som skal hjælpe med at vedligeholde viden om sundhedsdata.

20.3 Konklusion

Ovenstående afsnit illustrerer, at en implementering af den nationale biobank og dermed dette systemkoncept vil medføre konsekvenser for SST og SSI set ud fra et arbejdsdelings- og økonomisk perspektiv.

Overordnet vil implementeringen medføre flest konsekvenser for SST, idet Forskerservice vil miste nogle funktioner og dermed også økonomisk indtægt. Implementeringen vil også have konsekvenser for SSI, idet de risikerer at forskere, der udelukkende ønsker dataudtræk, vil kontakte dem, selvom dette eventuelt ikke var hensigten med den nationale biobank. Ligeledes skal de varetage flere opgaver, og implementeringen vil give konstante drift- og personaleomkostninger for SSI.

Til trods for disse konsekvenser for SST og SSI vurderes det, at det valgte kompromis for datahåndteringen er hensigtsmæssig at indføre, hvis der opstilles nogle restriktioner for henvendelse ved dataudtræk. Denne vurdering bygger på, at SST kun vil miste et fåtal af funktioner, samtidig med at de fortsat vil være registeransvarlige. Ydermere skal SSI ikke varetage flere opgaver end forventet og de økonomiske konsekvenserne i forbindelse med drift og personale vil ligeledes ikke være større end forventet.

Del VI

SYNTESE

I denne del diskuteres Løsningsstrategien samt metode til koncepttest og koncepttestresultater. Ligeledes konkluderes der på problemformuleringen, og der foretages en perspektivering på projektet.

Diskussion

21.1 Evaluering

21.1.1 Resultater

Resultaterne af evalueringen demonstrerede, at SST gerne vil bidrage til en løsning for datahåndteringen i den nationale biobank. Dog ønsker de ikke, at andre skal håndtere deres data, hvorfor de appellerer til en løsning, der tillader at de kan være egenrådige over deres data. På baggrund af dette kan det diskuteres om en implementering af den nationale biobank, inklusive Mads Melbyes visioner, overhovedet er mulig, da enighed om datahåndteringen er nødvendigt.

Det blev ligeledes klart i evalueringen, at den opstillede systemarkitektur er meget lig planerne for den nationale biobank, hvilket var forventeligt set ud fra visionerne for den nationale biobank og korrespondencerne med Mads Melbye og Kristian Hveem foretaget i løbet af projektperioden.

Endvidere viste koncepttestresultaterne som forventet, at de anvendte standardmetoder overordnet er anvendelige. Dog er der mulighed for at foretage nogle ændringer, der vil forbedre systemet. Selve opdateringsfunktionen, hvor alle sundhedsdata i Biobankregisteret ryddes og alle relevante sundhedsdata indsættes, er en uhensigtsmæssig løsning, idet det bevirker, at alle CPR-numre tildeles nye ID ved hver opdatering. Hermed kan forskere ikke koble tidligere data med nye data, idet den samme patient vil have flere forskellige ID.

Da en del af Løsningsstrategien var at skitsere, hvorledes forskere vil få data præsenteret, skulle en fuldstændig koncepttest have inkluderet forskeres vurdering mht. datapræsentation. Derfor vil det være nødvendigt at kontakte disse for at få en vurdering af, hvorvidt skærmbilledernes design, brugervenlighed samt funktionaliteter er optimal.

21.1.2 Metode

Det kan endvidere diskuteres, hvorvidt evalueringen af det designede systemkoncept kunne være foretaget anderledes. Inkludering af flere respondenter, såsom flere medarbejdere fra SST og SSI, epidemiologer og systemudviklere, i evalueringen, kunne have bidraget med et mere endegyldigt billede af, hvorvidt det valgte kompromis samt datahåndteringssystemet er en mulig løsning på datahåndteringen i den nationale biobank. Ydermere er det forventeligt, at andre resultater ville være gældende, hvis koncepttesten var blevet foretaget løbende, idet der på et tidligt tidspunkt kunne opnås indsigt i, hvorvidt de indblandede parter fandt kompromiset samt datahåndteringen hensigtsmæssig. Det vurderes dog, at evaluering foretaget i slutningen af projektets udfærdigelse er tilstrækkelig, idet gruppen har haft indgående kendskab til visionerne med den nationale biobank, samt SST's holdning til datahåndteringen fra begyndelsen af projektet.

21.2 Konsekvensvurdering

Før datahåndteringssystemet implementeres ud fra det valgte kompromis, bør konsekvenserne, som en implementering vil medføre, tages i betragtning. Konsekvenserne består i at SST's Forskerservice vil miste nogle funktioner, hvorfor det ligeledes vil have nogle økonomiske konsekvenser for SST. Når implementeringen er foretaget vil SSI have brug for vedligeholdelse af kendskab til deres data, og derfor er det oplagt, at SST skal stille ekspertressourcer til rådighed for SSI. På baggrund af dette kan det diskuteres, hvorvidt SST ønsker at være SSI behjælpelig, netop fordi SST mister nogle funktioner samt indtægter ved denne løsning. Det er derfor nærliggende at tro, at SST skal have nogle helt klare definitioner på, hvad gevinsten er for dem ved etablering af den nationale biobank. Er disse ikke til stede, skal de eventuelt identificeres, hvorfor første led i denne proces kan være opstillingen af restriktioner for dataudtræk, som beskrevet i afsnit 20.1 på side 92.

21.3 Systemudviklingsmetode

Udviklingen af dette systemkoncept er foretaget med inspiration fra UP, som er en velegnet metode i forhold til systemudvikling samt konsekvensvurdering. Denne metode har været værdifuld, idet det er en fleksibel metode, der tillader tilføjelser samt revurderinger i løbet af systemudviklingen. Ved implementering af systemet kan det dog overvejes, hvorvidt en anden metode kunne gøre implementeringen lettere. En mulighed er eksempelvis at anvende Scrum-metoden, der er en metode der anvender iterativ udvikling til hurtigt at udvikle og demonstrere brugbar software hver anden til fjerde uge (ATIVE, 2010).

21.4 Kompromis

Det kan ligeledes diskuteres om valget af dette kompromis, er det mest hensigtsmæssige i fht. at løse problemet med de modsatrettede holdninger om håndtering af data. De to yderligere kompromiser, der blev identificeret i Løsningsstrategien vil være vanskelige at indføre, idet både SST og SSI formentlig ikke vil afgive hhv. data og prøver. En anden mulighed er, at sygehusene indberetter data til SST på sædvanlig vis, samtidig med at de ligeledes indberetter relevante sundhedsdata direkte til SSI. Det er dog uvist, om dette kan indføres, idet det ikke er sikkert at sygehusene vil indberette data medmindre dette er lovpligtigt.

Det valgte kompromis fokuserer som beskrevet på, at SST ikke må miste deres data og SSI skal besidde relevante sundhedsdata fra SST, hvorfor kopiering af relevante sundhedsdata var en oplagt metode at anvende. Kopiering af data fra SST til SSI betyder dog, at de samme data eksisterer to steder, hvilket kan være uhensigtsmæssigt, da der kan opstå tvivl om hvilke data der er gældende. Et andet perspektiv er derfor, at alle sundhedsdata (journaldata, registerdata osv.) i hele sundhedssektoren og alle biologiske prøver landet over samles i én fælles data/biobank. Dermed sikres det, at alle data og prøver kun findes ét sted, hvilket giver bedre overskuelighed for brugerne.

21.5 Adgang til relevante sundhedsdata

Ved implementering af Biobankregisteret og den nationale biobank kan det være fordelagtigt, at forskere får lettere adgang til rå sundhedsdata. En løsning kunne være at give forskere adgang til rå data fra den online webportal, således at der ikke skal modtages en CD-rom med dataudtræk. For at muliggøre dette skal den online webportal have links til såvel Datatilsynet som SSI. Når forskeren ønsker at få personhenførbare dataudtræk, kan der via det direkte link til Datatilsynet laves en online ansøgning om specifikke dataudtræk. Når denne er behandlet, vil forskeren modtage en godkendelse fra Datatilsynet via mail. Herefter kan forskeren anvende det direkte link til SSI, hvorfra godkendelsen samt ansøgningen til SSI kan foretages. Når SSI har behandlet disse, vil forskeren modtage en mail med et brugernavn og et kodeord, der giver adgang til præcis de ønskede sundhedsdata i Biobankregisteret. Derefter kan forskeren få visualiseret de rå sundhedsdata i eksempelvis et Excel ark. Det er uvist om denne process vil lette sagsgangen, idet der fortsat skal foretages ansøgning hos Datatilsynet og SSI for dataudtræk. Dette vil dog være i god tråd med den stigende digitalisering i sundhedssektoren, og der er mulighed for en tidsmæssig gevinst, idet alle ansøgninger og svar foregår online.

Konklusion

SSI tog i 2009 initiativ til at etablere en national biobank, der skal inkludere relevante sundhedsdata fra SST's registre og biologiske prøver fra danske biobanker, med det formål at forskere kan tilgå disse fra samme platform. Der er dog en uoverensstemmelse om datahåndtering da SST som registeransvarlig ikke ønsker at udlevere deres data til SSI, idet de mener, at de har den nødvendige ekspertviden om registerdata for at anvende dem. Derfor fokuserer denne rapport på at identificere en datahåndteringsarkitektur til understøttelse af et kompromis, hvor SST bibeholder deres funktion som registeransvarlig, samtidig med at SSI besidder en kopieret version af SST's relevante sundhedsdata. Ud fra dette var det centrale i denne rapport at skitsere en løsning til, hvorledes et datahåndteringssystem kan designes vha. standardmetoder, og samtidig understøtte det valgte kompromis.

Med inspiration fra Unified Process samt brug af Entity Relationship modellering blev der designet et system ved hjælp af standardmetoder, der tilgodeser SST's ønske om at beholde alle registerdata, samtidig med at SSI besidder de relevante sundhedsdata i Biobankregisteret. Use casen UC: "Kopier og gem data" er designet til at kopiere de relevante sundhedsdata fra SST's database og lagre de kopierede data i Biobankregisteret hos SSI. UC: "Visualiser data" er designet til at tillade forskeren at visualisere udvalgte aggregerede sundhedsdata. Datahåndteringssystemet består af to applikationer samt Biobankregisteret. Den ene applikation foretager kommunikationen med SST's database, opdateringen af Biobankregisteret, kopieringen og anonymiseringen af relevante sundhedsdata samt lagring af disse i Biobankregisteret. Den anden applikation tillader brugen af den online webportal og dermed den nationale biobank. Dermed er alle use cases opfyldt.

Evalueringen af systemet belyste, at SST er enige med SSI i, at den nationale biobank er et fornuftigt og gavnligt koncept. Ligeledes kan SST se det fornuftige i, at kompromiset lægger op til en arbejdsdeling mellem SST og SSI, således at alle kernekompetencer udnyttes fuldt ud. Dog er SST ikke enige i, at SSI er nødsaget til at besidde de relevante sundhedsdata for at lave en brugbar løsning til forskere, idet SSI ikke har de fornødne

kompetencer til at anvende SST's data. Endvidere gav Mads Melbye fra SSI udtryk for, at den skitserede datahåndteringsarkitektur stemmer godt overens med SSI's planer for datahåndteringen. Evalueringen af datahåndteringssystemets tekniske forhold demonstrerede, at standardmetoder var anvendelige til udførelse af systemkonceptet, omend noget pragmatiske.

Det valgte kompromis samt datahåndteringssystem vil ligeledes have nogle konsekvenser for SST og SSI efter implementering. Efter etablering af den nationale biobank risikerer SST's Forskerservice at miste nogle funktioner, idet forskere vil rette henvendelse til SSI, hvis det ønskes at koble biologiske prøver og sundhedsdata. Dog mister SST ikke deres funktion som registeransvarlig, idet Biobankregisteret udelukkende vil indeholde kopierede relevante sundhedsdata, og SST skal derfor stadig varetage kontakten til sygehusene og styre indberetningen af data. Ydermere vil driften af datahåndteringssystemet kræve indgående viden om data, hvorfor SSI vil få brug for SST's ekspertviden om disse. Yderligere vil der være nogle økonomiske konsekvenser for både SST og SSI. SST vil miste indkomst fra Forskerservice, idet nogle funktioner ikke længere skal varetages af dem, hvorimod SSI vil have vedvarende økonomiske driftomkostninger mht. personalelønninger og teknik.

Det skitserede og designede datahåndteringssystem viser, at det er muligt at lave et system vha. standardmetoder, hvor SST fortsat er registeransvarlige samtidig med at SSI besidder de relevante sundhedsdata, der er nødvendige for at den nationale biobank kan etableres efter hensigten. Dog vil implementeringen af den nationale biobank og dermed datahåndteringssystemet medføre nogle arbejdsdelings- og økonomiske konsekvenser for både SST og SSI. Da SST ikke kan godkende valg af kopiering som metode til dataerhvervelse for SSI, skal der udvikles et andet datahåndteringssystem til understøttelse af kompromiset for datahåndteringen, der teknisk set er mulig at implementere, samtidig med at der er enighed om, hvordan data skal håndteres, så både SST og SSI tilgodeses.

Perspektivering

Det blev i forrige afsnit konkluderet at et datahåndteringssystemet, til understøttelse af det identificerede kompromis fra afsnit 5 på side 36, kan designes vha. standardmetoder, hvormed en implementering ud fra et teknisk synspunkt er muligt. Dermed er det ikke den tekniske løsning for den nationale biobank, der kan forhindre en udførelse af biobanken i henhold til de opstillede visioner. I stedet bør fokus være på den manglende enighed om datahåndteringen der udspringer af at der skal ske fundamentale ændringer i fht. de nuværende forhold. SST har ikke nogen interesse i, at andre skal varetage deres funktioner og få deres data, hvorimod SSI har gang i et projekt der er af stor betydning for dem som forskningsinstitut, og begge instanser er af stor betydning i det danske samfund, hvorfor en løsning ikke kan presses igennem. Derfor kan det virke umuligt at et kompromis der tillader adgang til sundhedsdata og biologiske prøver i samme arbejdsgang kan findes.

Derfor kunne der også anvendes en anden tilgang til problemstillingen som inkluderede at få SST og SSI til at tale sammen om deres forventninger til hvordan den nationale biobank kan udføres så den opfylder visionerne bedst muligt. Hermed kunne SST fortælle om de betænkningsde har ved at lade andre håndtere deres sundhedsdata, og SSI kunne fremlægge deres synspunkter for, hvordan datahåndteringen kunne foretages. Dermed kunne de to parter muligvis komme et trin nærmere et kompromis der kan accepteres af dem begge.

LITTERATURLISTE

- ATIVE (2010). Implementering af scrum. <http://www.ative.dk/media/66/scrum.pdf>.
- Beynon-Davies, P. (2000a). *Database systems*, kapitel 7, side 359-365. Palgrave, 2. udgave. ISBN 0-333-79227-0.
- Beynon-Davies, P. (2000b). *Database systems*, kapitel 7, side 406. Palgrave, 2. udgave. ISBN 0-333-79227-0.
- Det Centrale Personregister (2010). Folkeregistrering. <http://www.cpr.dk/cpr/site.aspx?p=308>.
- Dillner, J. (2002). Unik chans till förbättring. *Läkartidningen*, 99:2274-2276.
- Elmasri, R. og Navathe, S. B. (2003a). *Fundamentals of Database Systems*, kapitel 1, side 15-20815. Pearson, 4. udgave. ISBN 0-321-20448-4.
- Elmasri, R. og Navathe, S. B. (2003b). *Fundamentals of Database Systems*, kapitel 25, side 812-815. Pearson, 4. udgave. ISBN 0-321-20448-4.
- Elmasri, R. og Navathe, S. B. (2003c). *Fundamentals of Database Systems*, kapitel 9, side 279-280. Pearson, 4. udgave. ISBN 0-321-20448-4.
- Elmasri, R. og Navathe, S. B. (2003d). *Fundamentals of Database Systems*, kapitel 8, side 247. Pearson, 4. udgave. ISBN 0-321-20448-4.
- Elmasri, R. og Navathe, S. B. (2003e). *Fundamentals of Database Systems*, kapitel 8, side 246. Pearson, 4. udgave. ISBN 0-321-20448-4.
- Elmasri, R. og Navathe, S. B. (2003f). *Fundamentals of Database Systems*, kapitel 8, side 245-247. Pearson, 4. udgave. ISBN 0-321-20448-4.
- Eriksson, H.-E., Penker, M., Lyons, B. og Fado, D. (2004a). *UML 2 Toolkit*, kapitel 1, side 1-15. Wiley Publishing. ISBN 0-471-46361-2.

- Eriksson, H.-E., Penker, M., Lyons, B. og Fado, D. (2004b). *UML 2 Toolkit*, kapitel 5, side 145-191. Wiley Publishing. ISBN 0-471-46361-2.
- Eriksson, H.-E., Penker, M., Lyons, B. og Fado, D. (2004c). *UML 2 Toolkit*, kapitel 2, side 24-36. Wiley Publishing. ISBN 0-471-46361-2.
- Folketinget (2010). Sammenskrevet udgave af persondataloven. <http://www.datatilsynet.dk/lovgivning/persondataloven>.
- HUNT Biosciences AS (2008). Non-confidential information.
- Jonsson, L. og Landegren, U. (2004). Storing and using biobanks for research. *Rudbeck Laboratory, Department of Genetics and Pathology, Uppsala University*, side 1-10.
- Kommisionen for de europæiske fællesskaber (2008). Sjældne sygdomme: en udfordring for europa. http://ec.europa.eu/health/ph_threats/non_com/docs/rare_com_da.pdf.
- Landegren, U. (2002). Genteknologier och biobanker förklarar sjukdomsmekanismer. *Läkartidningen*, 99:1920-1921.
- MedCom (2010). Snomed-koder. <http://www.medcom.dk/wm109881>.
- MedCom, Indenrigsministeriet og Sundhedsministeriet (2008). *Den Gode Webservice*, side 18-20. MedCom, 11. udgave.
- MedCom og Sundhedsministeriet, I. (2008a). *Den Gode Webservice*, side 1-9. MedCom, 11. udgave.
- MedCom og Sundhedsministeriet, I. (2008b). *Den Gode Webservice*, side 30-36. MedCom, 11. udgave.
- Ministeriet for Sundhed og Forebyggelse, Sundhedsstyrelsen, Danske Regioner og Finansministeriet (2009). Bedre dokumentation.
- Net 2000 Ltd. (2005). Data sanitization techniques, a net 2000 ltd. white paper.
- Novo Nordisk Fonden (2009). Årsskrift 2008/2009.
- Pfleeger, S. L. (2001). *Software Engineering Theory and Practice*, kapitel 4, side 141-142. Alan R. Apt, 2. udgave. ISBN 0-13-029049-1.
- Ponniah, P. (2001). *Data Warehousing Fundamentals*, kapitel 7, side 127-141. John Wiley and Sons, 1. udgave. ISBN 0-471-41254-6.
- Scott, K. (2001). *The Unified Process Explained*. Pearson Education, 1. udgave. ISBN 0-201-74204-7.

- Silberschatz, Korth og Sudarshan (2001a). *Database System Concepts*, kapitel 2, side 36-65. McGraw-Hill, 4. udgave.
- Silberschatz, Korth og Sudarshan (2001b). *Database System Concepts*, kapitel 3, side 90-93. McGraw-Hill, 4. udgave.
- Silberschatz, Korth og Sudarshan (2001c). *Database System Concepts*, kapitel 2, side 71-75. McGraw-Hill, 4. udgave.
- Silberschatz, Korth og Sudarshan (2001d). *Database System Concepts*, kapitel 7, side 260-290. McGraw-Hill, 4. udgave.
- Statens Institut for folkesundhed (2007). Sundhedsregistre og befolkningsundersøgelser i danmark - til brug ved planlægning af borgerrettet forebyggelse i kommunen.
- Sundhedsstyrelsen (2010a). Dokumentation af dødsårsagsregisteret.
- Sundhedsstyrelsen (2010b). Dokumentation af fødselsregisteret.
- Sundhedsstyrelsen (2010c). Dokumentation af landspatientregisteret.
- Sundhedsstyrelsen (2010d). Dokumentation af patologiregisteret.
- Sundhedsstyrelsen (2010e). Forskerservice. <http://www.sst.dk/Indberetning%20og%20statistik/Forskerservice.aspx>.
- Sundhedsstyrelsen (2010f). Hvad er sks. http://www.sst.dk/Indberetning%20og%20statistik/Klassifikationer/Hvad_er_SKS.aspx.
- SWEGENE (2003). The swedish national biobank program. <http://www.biobanks.se/>.
- UK Biobank Coordinating Centre (2006). Uk biobank: Protocol for a large scale prospective epidemiological resource.
- Umanath, N. S. og Scamell, R. W. (2007). *Data Modelling and Database Design*, kapitel 1, side 11-13. Thomson Course Technology, 1. udgave. ISBN 1-4239-0083-9.
- Videnskabsministeriet (2009). Danmark får national biobank i verdensklass. *Pressemeddelelser*, side 1-2.

APPENDIKS

Interviewguide

Dataindsamlingen er foretaget via syv kvalitative informantinterviews og tre kvalitative, semi-strukturerede interview. Både informantinterviews og semi-strukturerede interviews skulle primært bidrage til besvarelse af den initierende problemstilling. Valget af kvalitative semi-strukturerede interviews som metode har givet mulighed for uddybninger og specificeringer omkring visionerne, status og kendskab til den nationale biobank, mens de kvalitative informantinterviews har bidraget til at skabe en forståelse for registrerede sundhedsdata samt registerforhold. De adspurgte respondenter anses, på baggrund af deres profession, for at være valide respondenter i forhold til besvarelsen af de stillede spørgsmål. De syv informantinterviews er af ca. fem minutters varighed, hvorimod de semi-strukturerede interviews er af ca. 20 minutters varighed, idet interviewet er semi-struktureret og flere spørgsmål skal besvares. Alle interviews vil blive optaget med henblik på efterfølgende transskribering.

I forbindelse med de semi-strukturerede interviews, er et kvalitativt undersøgelsesdesign anvendt, hvor de to personer fra SSI, der er ansat til at styre projektet med den nationale biobank, samt én person fra SST der er medarbejder i SST's dokumentationsenhed, vil blive interviewet. Den kvalitative metodes udgangspunkt er den interviewedes subjektivitet, og endvidere at dataindsamlingen skal bidrage til dybere forståelse af helheden. Fordelen ved dette undersøgelsesdesign er bl.a. muligheden for at gå i dybden med problemstillingen, hvor man spørger ind til respondentens holdninger.

A.1 Semi-strukturerede interviews

A.1.1 Mads Melbye

- Vi har læst dine visioner, som er skrevet til Novo Nordisk Fonden, hvor der står at data bl.a. skal opsamles fra Landspatientregisteret og Patologiregisteret. I den forbindelse vil vi høre om det stadig pt. er en del af visionerne med den nationale

biobank?

- Hvor langt er I i forhold til at kunne realisere og kunne hente de her data?
- Så vidt vi kan se, så skal I have fat i Sundhedsstyrelsen for at få fat i de forskellige data som I ønsker
- I snakker i visionerne om en on-line webportal, hvor forskere kan få direkte adgang til data. Har I gjort jer nogle overvejelser om, hvordan I vil håndtere anonymiseringen af de her data?
- Så det er altså også for at forskere kan få et overblik over, hvorvidt der eksisterer biologisk materiale inden for det her område?
- Dvs når det her projekt kommer op at køre, så vil forskere skulle gøre det samme som de pt. gør hos Sundhedsstyrelsen, når de ønsker adgang til data, hvor en ansøgning sendes til Forskerservice?

A.1.2 Kristian Hveem

- Hvor langt er I kommet med projektet med den nationale biobank?
- De data I gerne vil have ind i Biobankregisteret om patienterne, og de biologiske prøver der skal ind i systemet, hvordan skal de opsamles?
- Er det meningen at biobanken i den færdige stand skal bruges af læger på sygehusene eller primært til forskning i første omgang?
- Er der nogen steder skrevet ned hvilke visioner og strategier der er for biobanken?
- Er det noget vi kan finde et sted eller er det muligt at det er noget vi kan se?
- Vi har forstået, at du er direktør for en biobank i Norge. Bliver biobanken i Danmark noget lignende den der er i Norge?
- Det vil sige på Rigshospitalet har de alle de patologiske prøver i deres eget system, og de vil gerne beholde det?

A.1.3 Sundhedsstyrelsen

- Kender du til projektet om den nationale biobank i Danmark?
- Hvis ja, hvad bliver sundhedsstyrelsens funktion indenfor registrering af sundhedsdata, når den nationale biobank er etableret?
- Hvis nej, forklares formålet med den nationale biobank?

- Hvad er grundlaget for at data ikke skal lagres i den nationale biobank?
- Hvis der skal laves en online adgang er det så ikke lettere at have alle data et sted?
- Er det muligt at kopiere data i stedet for at flytte dem helt?

A.2 Informantinterviews

A.2.1 Onkologisk afdeling, Vejle Sygehus

Interview med lægesekretær på Onkologisk Afdeling, Vejle Sygehus.

- Lagres patientens diagnose?
- Lagres specifikke behandlingsvalg og medicinske undersøgelser?
- Lagres behandlingsoutcome? Et eksempel er, at en kræftpatient modtager behandling og denne fungerer, hvorfor tumoren mindskes. Lagres denne ændring så i PAS?
- Hvad så når tumoren er helt væk? Noteres det hvis patienten er erklaret rask selvom denne stadig skal gå til kontrol?
- Hvilke databaser indberetter I til?

A.2.2 Patologisk institut

Interview med lægesekretær fra Patologisk Institut, Skejby Sygehus.

- Bliver patologisvar fra undersøgelser lagret i sygehusets PAS system?
- Hvor lagres patologisvar så, hvis det lagres overhovedet?
- Udtrækker landspatientregisteret disse data?

A.2.3 Landspatientregisteret

Interview med informationsmedarbejder fra Sundhedsstyrelsen.

- Hvordan får landspatientregisteret kræftdata?
- Udtrækkes data om behandlingsvalg, diagnose, behandlingsoutcome?
- Hvor ofte opdateres Landspatientregisteret?
- Hvordan får forskere adgang til data fra Landspatientregisteret?

A.2.4 Fødselsregisteret

Interview med informationsmedarbejder fra Sundhedsstyrelsen.

Hvorfra udtrækkes data til Fødselsregisteret?

A.2.5 Familie-relations databasen

Interview med informationsmedarbejder hos Statens Serum Institut.

- Hvorfra udtrækkes data til Familie-relations databasen?
- Hvordan er disse data tilgængelige for forskere?

A.2.6 Forskerservice

Interview med specialist fra Forskerservice, Sundhedsstyrelsen.

- Hvordan får en forsker udtræk af data fra jeres registre?
- Hvis en forsker ønsker udtræk fra flere registre på én gang, hvordan laver I så søgninger i de forskellige registre?
- Bruger I SQL til dataudtræk?
- Når data skal leveres til forskeren, bliver data så sat sammen eller leveres de enkeltvis?
- Hvordan leveres data til forskerene?

A.2.7 Sundhedsstyrelsen

Interview foretaget d. 18.5.2010 med medarbejder fra Sundhedsstyrelsens IT-afdeling.

- Har hvert register sin egen database?
- Hvordan håndterer SST indberetningen fra PAS til SST's registre?
- Hvilke processer benyttes til at få data fra registre og ind i data warehouse?
- Vi ved at Landspatientregisteret opdateres hver nat, men ved du hvordan denne funktionalitet er sat op?

Semi-strukturerede interviews

Alle semi-strukturerede interviews blev udført telefonisk og indledt med følgende beskrivelse: Vi er to studerende fra Aalborg universitet, der skriver speciale om en fælles database i Danmark. I den forbindelse har vi nogle spørgsmål bl.a. angående visionerne for projektet med den nationale biobank og håndtering af sundhedsdata.

B.1 Mads Melbye

Interview foretaget d. 4.5.2010 med initiativtager og leder af den nationale biobank, Mads Melbye.

Vi har læst dine visioner, som er skrevet til Novo Nordisk Fonden, hvor der står at data bl.a. skal opsamles fra Landspatientregisteret og Patologiregisteret. I den forbindelse vil vi høre om det stadig pt. er en del af visionerne med den nationale biobank?

Visionerne er stadig at bygge en fysisk biobank, som bliver bygget op og som vil indeholde biologiskeprøver samt et Biobankregister, som vil indeholde information om de biologiske prøver der ligger i biobanken, hvor det så skal være muligt at udtrække hvordan det er gået de personer som ligger bag de biologiske prøver, så man som forsker kan få et overblik over hvad der findes potentielt set af materiale på personer der f.eks. har udviklet brystkræft.

Hvor langt er I i forhold til at kunne realisere og kunne hente de her data? Den del af projektet er vi slet ikke startet på endnu. Lige nu er vi kun i gang med at planlægge opbygningen af den store fysiske biobank. Det er sådan en lang process, hvor tingene skal i udbud, og der er en masse ting i forbindelse med det, så selve registerdelen, den har vi slet ikke rørt ved endnu.

Så vidt vi kan se, så skal I have fat i Sundhedsstyrelsen for at få fat i de forskellige data som I ønsker Ja det er ganske rigtigt og dem har vi også haft nogle indledende diskussioner med, men vi er ikke gået videre med den del af projektet. Det trækker virkelig tænder ud at få bygget sådan et byggeri og få det hele på plads, så det er det vi er startet med. Så

tager vi Biobankregisteret i anden fase.

I snakker i visionerne om en on-line webportal, hvor forskere kan få direkte adgang til data. Har i gjort jer nogle overvejelser om hvordan i vil håndtere anonymiseringen af de her data? Ja det er jo utrolig vigtigt, men der vil ikke i web-portalen blive adgang til noget der er bare i nærheden af en personidentificerbar oplysning. Det er aggregerede data som du får her. Så du får ikke individdata om du vil. Det bliver fuldstændig anonymiserede og så bliver det aggregerede data, det vil sige, det bliver mere i grupper eller f.eks. hvilke blodprøver er lavet i det og det tidsrum i Danmark og med hvilken diagnose, som du kan få at vide. Vil du vide mere så skal du til at søge, og så skal du den normale vej rundt i systemet. Lige præcis det med oplysninger, så har forskere længe efterlyst information om, hvorvidt der overhovedet biologisk materiale inden for det og det område.

Så det er altså også for at forskere kan få et overblik over hvorvidt der eksisterer biologisk materiale inden for det her område? Ja og jeg vil sige at Sverige også er igang med ettilsvarende system.

Dvs når det her projekt kommer op at køre, så vil forskere skulle gøre det samme som de pt. gør hos Sundhedsstyrelsen, når de ønsker adgang til data, hvor en ansøgning sendes til Forskerservice? Selve sagsgangen kan jeg først sige noget om når vi kommer tættere på slutningen af projektet, men det vil være sådan at de samme instanser, som også ansøges i dag, skal kontaktes igen i en eller anden forstand. Så det er der ikke noget nyt i. Som jeg sagde før så er man i Sverige og Norge også begyndt at diskutere det, så vi sidder faktisk og diskuterer en nordisk sammenhæng, og om det er muligt at lave et nordisk register eller en nordisk samling af data, således at norden kan udnytte at vi ikke kun har verdens bedste register men også de biologiske prøver, der skal sikre nordisk forskning i verdensklasse.

B.2 Kristian Hveem

Interview foretaget d. 16.2.2010 med direktør af den nationale biobank, Kristian Hveem.

Hvor langt er i kommet med projektet med den nationale biobank? Vi er ikke kommet så langt med projektet. Først og fremmest er der tale om et byggeri til de mange biologiske prøver, der indledningsvist skal arbejdes på at blive bygget. Derfor skal der tages højde for mange millioner histologiske prøver, som skal udtages og data der beskriver disse, skal placeres i et nyt datainformationssystem.

De data i gerne vil have ind i Biobankregisteret om patienterne, og de biologiske prøver der skal ind i systemet, hvordan skal de opsamles? Foreløbig skal vi ikke have vævsprøver.

Så det er data omhandlende diagnose og behandling I ønsker at opsamle? Der vil kunne dannes en patienthistorik, men først og fremmest er der tale om en kobling til andre registre.

Så det bliver en opkobling til allerede eksisterende registre, f.eks. Landspatientregisteret, således at data kan udtrækkes? Ja.

Er det meningen at biobanken i den færdige stand skal bruges af læger på sygehusene eller primært til forskning i første omgang? Det er en forskningsbiobank.

Er der nogen steder skrevet ned hvilke visioner og strategier der er for biobanken? Ja, i ansøgningen om bevilling til projektet, og Mads Melbye har skrevet en artikel om strategierne for den nationale biobank.

Er det noget vi kan finde et sted eller er det muligt at det er noget vi kan se? Jeg vil spørge Mads Melbye om det er okay, og så sender jeg dem til jer.

Vi har forstået at du er direktør for en biobank i Norge. Bliver biobanken i Danmark noget lignende den der er i Norge? Det bliver ganske ens bortset fra, at de har bygget biobanken op om en helseundersøgelse for en stor population som bor i en bestemt region. Den nationale biobank her vil først og fremmest være en biobank med prøver i forbindelse med vaccination og diabetes med henblik på at få en komplet diabetesbank. Dog skal Rigshospitalet skal have sine egne DNA-prøver i egen biobank.

Det vil sige på Rigshospitalet har de de alle de patologiske prøver i deres eget system og de vil gerne beholde det? Ja.

B.3 Sundhedsstyrelsen

Interview foretaget d. 29.4.2010 med medarbejder fra Sundhedsstyrelsens Dokumentationssenhed.

Kender du til projektet om den nationale biobank i Danmark? Ja jeg kender godt til den nationale biobank i Danmark. Den varetages af statens serum institut og er blevet bevilget penge af staten, novo nordisk samt EU. Vi er også repræsentanter fra Sundhedsstyrelsen, som er med til møderne omkring de overordnede visioner med den nationale biobank.

Hvis ja, hvad bliver sundhedsstyrelsens funktion indenfor registrering af sundhedsdata, når den nationale biobank er etableret? Sundhedsstyrelsens funktion forbliver uændret. Det er nemlig ikke meningen af den nationale biobank skal besidde alle data, men derimod data omhandlende patologi og genetik. Hvis den nationale biobank så ikke er i besiddelse af data, som er relevante i forhold til forskerens formål, kan Statens Serum Institut kontakte Sundhedsstyrelsen og få de relevante data udtrukket til forskeren. Men hvis det er sådan, at man har brug for noget biologisk materiale eller noget fra biobanken så, for at man ikke skal rende spidsrod mellem alle mulige forskellige institutioner, vil man også kunne få nogen data ud fra vores registre i forbindelse med at man skal have data ud fra serum instituttet.

Hvis nej, forklares formålet med den nationale biobank? ...

Kræver det så ikke stadigvæk, at dem som ansøger om data laver en ansøgning til jer om at få yderligere data end de data som den nationale biobank har? Ja nu skal vi finde ud af hvad sagsgangen skal være. De er jo også nødt til at have de fornødne tilladelser fra

datatilsynet for overhovedet at få data, derefter så er det jo vores ambitioner at gøre det så smidigt som muligt.

Hvori består fordelene ved den nationale biobank? Fordelen er at forskere hurtigere kan få adgang til anonymiserede data via f.eks. en online web-portal, og dermed kan forskeren undgå at kontakte sundhedsstyrelse og vente på at få adgang til data. Men jeg vil gerne understrege, at de visioner der i sin tid blev nedfældet om den nationale biobank, er noget anderledes og mere omfangsrige sammenlignet med det formål der er i dag.

Hvad er grundlaget for at data ikke skal lagres i den nationale biobank? Hvorfor skulle de det? Hvorfor skulle man gøre det? Det er jo spørgsmålet om, hvem der har ansvaret for hvilke registre, ellers var der i stedet et sted der skulle have ansvaret for alle registre i hele Danmark. Ellers ville du jo skabe en ny snitflade, og der er snitflader. Vi har forskellige kompetencer, forskellig viden om forskellige registerområder, og vi har derfor viden om landspatientregisteret, og dermed viden om hvad for nogen data der er i landspatientregisteret, hvis man gerne vil have informationer om det og det og det, og hvad det så er for nogen data man skal have. Det har vi en kæmpe viden om her i Sundhedsstyrelsen, fordi det er os der laver alle definitionerne i landspatientregisteret, det er os der har kommunikationen med sygehusene om hvad for nogen data de indrapporterer. Så det er derfor, at data ligesom skulle ud herfra, netop fordi vi kan give den der kompetente rådgivning ifht hvad landspatientregisteret er for et register. Men dermed siger vi så at, hvis folk er fuldkommen målrettede og de ved lige nøjagtig hvad for nogen data de vil have, så skal de bare have dem et sted, så er der ingen grund til at de skal søge og løbe rundt til alle mulige stationer for at få alle mulige data, så vil vi jo gerne give dem så god service som mulig.

Hvis der skal laves en online adgang er det så ikke lettere at have alle data et sted Det er også rigtigt, men man skal overveje hvad det er man har online adgang til, ellers skal datatilsynet inddrages. Og man er også nødt til at sikre sig at udtrækkene bliver sådan, at de bliver reelt anonymiserede så det ikke er muligt at identificere en patient. Så der er lidt udfordringer i at etablere sådan et system.

Kunne man kopiere data i stedet for at flytte dem helt? Jo men det kommer an på hvad for et fokus du har, og har du fokus på at producere data og have viden om data, at opretholde kvaliteten af data eller har du fokus på at forskere skal gå et sted hen og hente alle sine data. Og det vi har lavet med Forskerservice er jo et forsøg på at sige, at så er der en indgang, og så kan man henvende sig et sted for at få alle de data man skal have på sundhedsområdet. Det der så bare sker er, at der er nogle registre der ligger hos sundhedsstyrelsen og der er nogle registre der ligger hos statens serum institut, og de registre har vi ikke adgang til. Det handler om hvordan reglerne sat op. Vi har ikke adgang til lægemidler og derfor kan vi ikke udlevere lægemiddelstatistik og på samme måde har lægemiddelstyrelsen ikke adgang til vores registre. Der vil komme nogen andre views og konflikter med det her. Hvorfor er det Biobankregisteret, der skal være omdrejningspunktet for at man skal udlevere alt

sundhedsregisterets data? Der er jo også hensyn til hvorfor lægemiddelregisteret ligger hos lægemiddelstyrelsen.

Men Statens Serum Institut er bare ikke gearet til at udlevere data fra sundhedsregisteret. En ting er, at de skal have biobanken op at stå, en anden ting er, hvad skal der kobles på af yderligere services, og hvordan skal de blive i stand til at levere de services. For det er også en udfordring at opretholde forskerservice. Det er ikke sådan helt klart for os allesammen, hvordan det ender med biobanken. Det er et kæmpe og godt projekt, men vi er ikke helt færdige, det handler ikke om sektor, men derimod om hvordan får vi skabt klarhed og sikkerhed. At kopiere data er altid farligt. Det er altid farligt, at have alt for mange kopier af data. Vores fokus er at vi gerne vil servicere forskere bedst muligt.

Informantinterviews

Alle informantinterviews blev udført telefonisk og indledt med følgende beskrivelse: Vi er to studerende fra Aalborg universitet, der skriver speciale om en fælles database i Danmark. I den forbindelse har vi nogle spørgsmål angående lagring af sundhedsdata. Vi ved ikke om vi har fået fat i den rette eller om der er en anden vi kan stille vores spørgsmål?

C.1 Onkologisk afdeling, Vejle Sygehus

Interview med lægesekretær på Onkologisk Afdeling, Vejle Sygehus.

Lagres patientens diagnose? Diagnosen bliver lagret og status bliver ændret hver eneste gang patienten har været indlagt.

Lagres specifikke behandlingsvalg og medicinske undersøgelser? Behandlingsvalg og medicinske undersøgelser bliver specificeret og gemmes via det grønne system (PAS).

Lagres behandlingsoutcome? Et eksempel er, at en kræftpatient modtager behandling og denne fungerer, hvorfor tumoren mindskes. Lagres denne ændring så i PAS? Ja. Der sker det, at når patienten bliver indlagt registreres en diagnose i sygehusets PAS, og en behandling startes. Er behandlingsforløbet længerevarende, registreres diagnosen løbende sammen med behandlingen, så det er muligt at aflæse om patienten får det bedre eller bliver dårligere. Når forløbet afsluttes registreres dette med en kode samt den sidste diagnose der er givet patienten. Dør patienten registreres dette med en kode, der indberettes til Dødsårsagsregisteret.

Hvad så når tumoren er helt væk? Noteres det hvis patienten er erklæret rask selvom denne stadig skal gå til kontrol? Ja, så registreres diagnosen med en kontrol-kode indtil forløbet er afsluttet og patienten er erklæret fuldstændig rask.

Hvilke databaser indberetter I til? Landspatientregisteret og Cancerregisteret (vha de koder der indtastes på patienterne).

C.2 Patologisk institut

Interview med lægesekretær fra Patologisk Institut, Skejby Sygehus.

Bliver patologisvar fra undersøgelser lagret i sygehusets PAS system? Nej det gør det ikke.

Hvor lagres patologisvar så, hvis det lagres overhovedet? Det bliver gemt i en patologisk database, som derefter indberettes til Patologiregisteret.

Udtrækker landspatientregisteret disse data? Det ved jeg ikke noget om.

C.3 Landspatientregisteret

Interview med informationsmedarbejder fra Sundhedsstyrelsen.

Hvordan får landspatientregisteret kræftdata? De registreres ude på sygehusene ved forskellige leverandører. LPR opdateres dagligt om natten. Private bruger sundhedsstyrelsens hjemmeside til at opdatere LPR med deres data. To indberetningsmuligheder (privat/ offentlig).

Udtrækkes data om behandlingsvalg, diagnose, behandlingsoutcome? Ja det bliver alt sammen udtrukket.

Hvor ofte opdateres Landspatientregisteret? Den opdateres én gang i døgnet, om natten.

Hvordan får forskere adgang til data fra Landspatientregisteret? De skriver en ansøgning til Forskerservice som derefter bliver behandlet. Det kan tage op til 30 dage at udtrække data og det koster ca. 1500 i timen.

C.4 Fødselsregisteret

Interview med informationsmedarbejder fra Sundhedsstyrelsen.

Hvorfra udtrækkes data til Fødselsregisteret? De udtrækkes primært fra Landspatientregisteret pånær ved hjemmefødsler, hvor der indsendes en blanket til Fødselsregisteret, hvorefter der udtrækkes data fra CPR-registeret.

C.5 Familie-relations databasen

Interview med informationsmedarbejder hos Statens Serum Institut.

Hvorfra udtrækkes data til Familie-relations databasen? Data udtrækkes udelukkende fra CPR-registeret.

Hvordan er disse data tilgængelige for forskere? De skal kontakte Professor Mads Melbye.

C.6 Forskerservice

Interview med specialist fra Forskerservice, Sundhedsstyrelsen.

Hvordan får en forsker udtræk af data fra jeres registre? Forskeren laver en ansøgning til datatilsynet, hvis der ønskes indsigt i personhenførbare data, og dernæst skal der ligeledes ansøges hos os. Ellers laver han en ansøgning til os om formål med forskningen samt specificering af de ønskede data.

Hvis en forsker ønsker udtræk fra flere registre på én gang, hvordan laver I så søgninger i de forskellige registre? Data fra f.eks. Landspatientregistret ligger i en række forskellige tabeller, og så specificerer vi vha. kode, hvilke tabeller og variabler der skal hentes ud, og hvilke kriterier der skal gælde som f.eks. alder, diagnose og behandling.

Bruger I SQL til dataudtræk? Ja

Når data skal leveres til forskeren, bliver data så sat sammen eller leveres de enkeltvis? Som hovedregel leveres data fra forskellige registre enkeltvis dvs. i de tabeller som de også er i i registre.

Hvordan leveres data til forskerne? Data leveres på en CD-rom og sendes anbefalet med et brev.

C.7 Sundhedsstyrelsen

Interview foretaget d. 18.5.2010 med medarbejder fra Sundhedsstyrelsens IT-afdeling.

Har hvert register sin egen database? Nej, der er et indberetningssystem knyttet til hvert register, som hver især lægger data over i en midlertidig database, og disse databaser er styret af forskellige DBMS. Herefter samler vi alt data i et data warehouse, der befinder sig i et DB2-system, således at data er i en og samme database. Og det er så herfra data kan tilgås.

Hvordan håndterer SST indberetningen fra PAS til SST's registre? Hospitalet/regionen hiver de krævede data ud af lokale PAS og/eller EPJ-systemer og indsender dem til os via FPT over sundhedsdatanettet i et bestemt format.

Hvilke processer benyttes til at få data fra registre og ind i data warehouset? Der er lavet en Java-applikation, hvori SQL-kommandoer benyttes til at lave udtræk fra de forskellige registre databaser og ind i data warehouset.

Vi ved at Landspatientregisteret opdateres hver nat, men ved du hvordan denne funktionalitet er sat op? Der indberettes hele tiden, men der opdateres kun én gang dagligt om natten. Der anvendes en JBOSS, JRULES.

Unified Process

UP er en iterativ og inkrementel metode, der inddeles i fire faser: Forberedelse, Etablering, Konstruktion og Overdragelse, som alle udgøres af fem work flows: Krav, Analyse, Design, Implementering og Test. De enkelte faser deles op i en række iterationer der inkluderer de fem work flows i højere eller mindre grad afhængig af antallet af iterationer. (Scott, 2001)

Forberedelse

Den primære målsætning for Forberedelsesfasen er at etablere en case for det foreslåede system, der understøtter brugen af det. Forberedelsesfasen udgøres af følgende trin:

- Fastslå målet for systemet. Hvad skal systemet indeholde, og hvordan skal det interagere med dets omgivelser.
- Hvilke overordnede funktionelle og ikke-funktionelle krav skal der stilles.
- Opstille kandidat-arkitektur ud fra de initielle versioner af modeller for Krav, Analyse, Design og Test.
- Identificere de afgørende risici, som projektet vil støde på, og specificere hvordan de skal håndteres.
- Specificere værdier, der gør projektet fordelagtigt, set ud fra et økonomisk og finansielt perspektiv.

(Scott, 2001)

Etablering

I Etableringsfasen er det vigtigt at få specificeret de resterende funktionelle krav samt etablere et solidt arkitektonisk udgangspunkt, hvorfra videreudvikling kan foretages. De

primære trin i Etableringsfasen er som følger:

- Udvikle modeller, der adresserer størstedelen af de funktionelle krav som ikke blev håndteret under iterering af Forberedelsesfasen.
- Udvikle det arkitektoniske udgangspunkt, der inkluderer alle de vigtigste arkitektoniske elementer samt beskrive denne arkitektur.
- Rette fokus mod de signifikante risici, der kan medføre evt. forsinkelser på projektet.
- Specificere det kvalitetsniveau som systemet skal opnå i forhold til ikke-funktionelle krav såsom adgangsbegrænsning.
- Konkretisere forretningsplan og projektplan på baggrund af viden opnået i Forberedelsesfasen.

(Scott, 2001)

Konstruktion

Det primære mål i Konstruktionsfasen er at bygge en working version af systemet der kan leveres til beta kunder, mens der fortsat arbejdes på at mindske risici der kan true leveringen af det færdige produkt. Yderligere testes systemet. De vigtigste trin i denne fase er:

- Færdiggøre use case modellen. (Hermed er alle funktionelle krav i overensstemmelse og adresseret i systemet).
- Færdiggøre alle andre modeller: Analyse, Design, deployment, Implementering og Test.
- Modificere arkitekturen i den grad det er nødvendigt for at sikre, at det leverede system reflekterer den hensigt, der var specificeret i arkitecturebeskrivelsen.
- Monitorere kritiske risici (dem der truer systemets viability) og betydningsfulde risici (dem der kan medføre forstyrrelser i fht. budget, tidsplan eller begge) og løbende mindske dem.

(Scott, 2001)

Overdragelse

Overdragelsesfasens primære mål er at levere en fungerende version af systemet til beta kunder, samtidig med at defekter opdaget af disse kunder behandles på succesfuld vis. De primære trin i Overdragelsesfasen er som følger:

- Forberede udrulning af systemet, hvilket involverer forberedelse af sitet samt rådgivning til kunderne om nødvendige opgraderinger til hardware og systemets software.
- Forberede en fyldestgørende dokumentation som beta kunderne behøves.
- Rulle systemet ud.
- Justere systemet alt efter de forskellige kundemiljøer. (f.eks. Netscape Communicationer vs. Microsoft Internet Explorer)
- Indsamle information om fejl og lave korrektioner rettidigt.

Krav

Krav opstilles hovedsageligt i Forberedelsesfasen, men yderligere krav kan findes i Etableringsfasen. Kravspecifikationer skal være en beskrivelse af, hvad der ønskes af et system, og samtidig skal de være udførlige, målelige, testbare, relateret til identificerede behov eller muligheder og defineret på et detaljeniveau, der er tilstrækkelig for system design. Der findes både funktionelle og ikke-funktionelle krav. Funktionelle krav beskriver en interaktion mellem systemet og dets omgivelser, og identificerer de nødvendige opgaver, handlinger eller aktiviteter, der skal fuldføres. Ikke-funktionelle krav specificerer kriterier eller begrænsninger for løsningen og den måde systemet kan opbygges. En vigtig model i krav workflowet er use case diagrammer, der illustrerer use cases og de personer eller objekter, der interagerer med dem. (Scott, 2001)

Analyse

Analyse foretages i både Forberedelsesfasen og Etableringsfasen og formålet er at opnå en reel forståelse af kravene og bruge denne forståelse til at opbygge momentum i forhold til Design. I Analysen er der mulighed for at håndtere problemstillinger såsom fordelingen af objekter, overensstemmelser og præstation, og samtidig skiftes fokus fra brugeren til de værktøjer udvikleren skal bruge for at kunne udvikle systemet korrekt. En vigtig del af Analyse workflowet er at identificere analyseklasser. (Scott, 2001)

Design

Design foretages også i lille grad i Forberedelsesfasen, men er primært en del af Etableringsfasen. Det er en kreativ og iterativ proces, hvor problemer transformeres til løsninger, som kan opdeles i et konceptuelt og teknisk design. Et konceptuelt design er det første skridt, der fortæller præcist, hvad systemet skal gøre, og påbegyndes ofte i Analyse workflowet. Teknisk design oversætter det konceptuelle design til en højere detaljegrad og muliggør

forståelse af hvilken hardware og software, der er nødvendig for at løse problemet. I design workflowet er dannelsen af designklasser og deployment-modellen vigtige elementer. (Scott, 2001)

Implementering

Implementering er en del af Konstruktionsfasen, og formålet er at bygge en working version af systemet så det er klar til evaluering af betakunder. Det primære resultat af dette workflow er udarbejdelsen af implementeringsmodellen, der indeholder source kode og executables, der tilsammen udgør det implementerede system. I Implementering workflowet arbejdes der ligeledes videre på deployment-modellen, der blev påbegyndt i Design. (Scott, 2001)

Test

Test er en del af Konstruktionsfasen og er med til at sikre at systemet virker som specificeret i krav. Resultatet af dette workflow er en testmodel der indeholder de forskellige inputs i form af planer, cases og procedurer, men også forventet output, defekter og evaluering. Black-box tests er ofte udledt fra use cases, hvorimod white-box tests udledes fra use case realiseringsdesignet. (Scott, 2001)

APPENDIKS **E**

Dataformat

Hos SST er de relevante sundhedsdata enten specificeret i form af SKS-kode (Sundhedsvæsenets klassifikationssystem) eller SNOMED-kode (Systematized Nomenclature of Medicine).

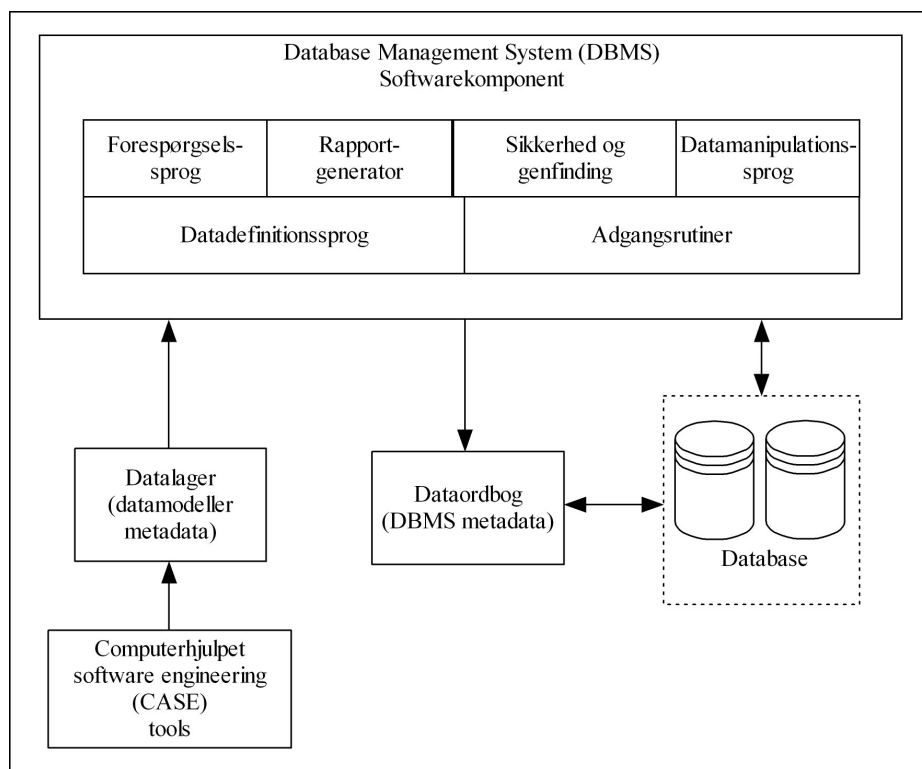
E.1 SKS og SNOMED

SKS er et klassifikationssystem som består af en samling af internationale, nordiske og danske klassifikationer, som anvendes af alle sundhedsfaglige personalegrupper, heriblandt forskere. SKS-koder danner grundlag for elektroniske informationssystemer til dokumentation og kommunikation inden for sundhedssektoren. (Sundhedsstyrelsen, 2010f) SNOMED er en anden klinisk sundhedsterminologi, der anvendes i forbindelse med svarafgivelse fra de patologiske afdelinger, hvor alle diagnoser er kodet ud fra SNOMED terminologien og tilhørende tekst (MedCom, 2010). Forskellen på SKS og SNOMED er, at SNOMED er den mest detaljerede og gennemarbejdede terminologi inden for det kliniske område. Ydermere er det hensigten, at den nyere version af SNOMED, også kaldet SNOMED CT, i fremtiden skal være den primære klassificeringsterminologi, og det vil derfor være oplagt, at alle sundhedsdata i fremtiden bliver beskrevet vha. SNOMED CT.

APPENDIKS F

DBMS

Figur F.1 illustrerer komponenterne i et databasesystem bestående af et DBMS, database, dataordbog og datalager.



Figur F.1: Komponenter i et Database system. (Umanath og Scamell, 2007)

DBMS er en samling af software, der er anvendelig til mange formål og som muliggør processerne til at definere, udforme og manipulere en database. De vigtigste komponenter i et DBMS inkluderer:

- et eller flere forespørgselsprog
- værktøjer til at generere rapporter

- faciliteter til at levere sikkerhed, integritetsbackup og recovery
- datamanipulationssprog for adgang til databasen
- et datadefineringssprog der bruges til at definere datastrukturen.

(Umanath og Scamell, 2007)

SQL har i relationelle DBMS en vigtig rolle i hvert af disse komponenter og bruges også i datadefinitionssproget (DDL) til at danne strukturen for databaseobjekter såsom tabeller, views og synonymer (f.eks. CREATE TABLE, DROP TABLE, ALTER TABLE osv.). Ligeledes genereres SQL-udsagn af programmeringssprog, der bruges til at bygge rapporter, for at få adgang til data fra databasen. Dataadministratorer bruger datakontrolsprog (DCL), der også gør brug af SQL-udsagn til både at kontrollere den "resource locking" der kræves i et multibruger miljø, facilitere backup og gendannelse fra driftssvigt samt sikre at brugere kun har tilgang til data som de er autoriseret til at bruge. (Umanath og Scamell, 2007)

Ligesom en ordbog er en referencebog, der giver information omkring ords form, oprindelse, mening og syntax, lagrer dataordbogen i et DBMS-miljø metadata, der giver information om definitioner af dataobjekter og deres relationer, autorisationer og brugerstatistikker. DBMS'et bruger dataordbogen til at slå de krævede datakomponentstrukturer og relationer op, og herved slipper applikationsudviklere (slutbruger og programmører) for at inkorporere datastrukturer og relationer i deres replikationer. Yderligere optages alle ændringer i databasens fysiske struktur automatisk i dataordbogen. Hermed fjernes behovet for at modificere de applikationsprogrammer, der tilgår den modificerede struktur. (Umanath og Scamell, 2007)

Datalageret er ikke en egentlig del af DBMS'et og vil derfor ikke blive behandlet her.

Korrespondence ang. Evaluering

G.1 Mads Melbye

Email

Kære Mads Melbye

Vi retter henvendelse til dig, da du tidligere har været behjælpelig i forbindelse med spørgsmål angående den nationale biobank, og vi håber derfor at du vil være behjælpelig endnu engang ved at besvare nogle spørgsmål.

I forbindelse med vores interview med dig angående den nationale biobank har vi erfaret at der indtil videre kun er fokus på den fysiske placering og opbygning af biobanken. Vi mener, at der skal være adgang til både specifikke registerdata samt biologiske prøver fra samme platform, hvis den nationale biobank skal have en gavnvirkning. I den forbindelse har vi kigget på, hvordan dette kan realiseres bedst muligt, således at det har færrest konsekvenser for både SSI og SST. Vi har derfor udarbejdet følgende konceptuelle løsning til hvordan Biobankregisteret kan etableres:

Vedhæftede figur illustrerer processerne for erhvervelse, lagring og præsentation af data i Biobankregisteret.

Vi forestiller os at de ønskede registerdata kan kopieres fra SST's registre, og lægges over i Biobankregisteret. Anonymiseringen af data foretages, idet vi forestiller os at sagsgangen ved udlevering af personhenførbare data evt. vil minde om sagsgangen hos SST, når de i dag udleverer data. Fra den online webportal skal der være mulighed for at få indsigt i aggregerede data samt antal patienter hvorpå der eksisterer biologiske prøver. Dermed kan forskere få udtræk af både rå data og aggregerede data, og det vil være muligt at koble viden om biologisk materiale med registerdata.

Spørgsmål:

- Kunne du forestille dig at ovenstående koncept kunne være en løsning for den na-

tionale biobank?

- Hvorfor?
- Hvordan forestiller du dig at en konceptuel løsning ellers kunne udfærdiges?

På forhånd tak for din hjælp

Med venlig hilsen Kate Z. Olesen og Ditte H. Kopp, 10. semester Sundhedsteknologi, Aalborg Universitet

Svar

Kære Ditte

Vi har ikke talt om dette men meget tæt på hvad du skitserer er hvad der ligger i planerne. Jeg vedlægger et skrift jeg skrev tidligere om disse ting. Ideen er at indhente få nøgleinfo fra de respektive biobanker til det nationale biobankregister. registeret trækker endvidere fra sundhedsstyrelsen informationer fra sundhedsregistrene etv således at du har den "komplete"info på de enkelte prøver. I aggregeret form lægges disse data på internettet således at forskere kan søge på hvormange og hvilke prøver der f.eks. eksisterer på sklerose patienter. Ved henvendelse til det koordinerende center kan man så få rådgivning om hvordan man søger (vid.etisk kommitee, etc.) for at få den reelle adgang til det biologiske materiale.

vh

Mads

G.2 Kristian Hveem

Email

Kære Kristian Hveem

Vi retter henvendelse til dig, da du tidligere har været behjælpelig i forbindelse med spørgsmål angående den nationale biobank, og vi håber derfor at du vil være behjælpelig endnu engang ved at besvare nogle spørgsmål.

I forbindelse med vores interview med dig angående den nationale biobank har vi erfaret at der indtil videre kun er fokus på den fysiske placering og opbygning af biobanken. Vi mener, at der skal være adgang til både specifikke registerdata samt biologiske prøver fra samme platform, hvis den nationale biobank skal have en gavnvirkning. I den forbindelse har vi kigget på, hvordan dette kan realiseres bedst muligt, således at det har færrest konsekvenser for både SSI og SST. Vi har derfor udarbejdet følgende konceptuelle løsning til hvordan Biobankregisteret kan etableres:

Vedhæftede figur illustrerer processerne for erhvervelse, lagring og præsentation af data i Biobankregisteret.

Vi forestiller os at de ønskede registerdata kan kopieres fra SST's registre, og lægges over i Biobankregisteret. Anonymiseringen af data foretages, idet vi forestiller os at sagsgangen ved udlevering af personhenførbare data evt. vil minde om sagsgangen hos SST, når de i dag udleverer data. Fra den online webportal skal der være mulighed for at få indsigt i aggregerede data samt antal patienter hvorpå der eksisterer biologiske prøver. Dermed kan forskere få udtræk af både rå data og aggregerede data, og det vil være muligt at koble viden om biologisk materiale med registerdata.

Spørgsmål:

- Kunne du forestille dig at ovenstående koncept kunne være en løsning for den nationale biobank?
- Hvorfor?
- Hvordan forestiller du dig at en konceptuel løsning ellers kunne udfærdiges?

På forhånd tak for din hjælp

Med venlig hilsen Kate Z. Olesen og Ditte H. Kopp, 10. semester Sundhedsteknologi, Aalborg Universitet

Svar

Vi har ikke modtaget noget svar fra Kristian Hveem, hvilket kan skyldes at Mads Melbye og Kristian Hveem har talt sammen indbyrdes om, hvem der skulle give svar.

G.3 Ansvarlig for SST's dokumentationsenhed

Email

Kære ...

Vi retter henvendelse til dig, da du tidligere har været behjælpelig i forbindelse med spørgsmål angående den nationale biobank, og vi håber derfor at du vil være behjælpelig endnu engang ved at besvare nogle spørgsmål.

I forbindelse med vores projekt om etableringen af den nationale biobank har vi igennem interview med dig erfaret, at der er uklarhed om brugen af SST's registerdata i den nationale biobank. Vi har i den forbindelse set nærmere på formålet med den nationale biobank, og forstået at det ønskes at give forskere let adgang til registerdata og biologiske prøver fra samme platform. Dette betyder, at SST's registerdata og de biologiske prøver skal bringes på samme platform, for at den nationale biobank skal have en egentlig gavnvirkning. Vi

har identificeret følgende tre kompromiser, hvormed SST's registerdata og de biologiske prøver kan bringes på samme platform:

1. Et kompromis kan bestå i, at de relevante data (Behandling, diagnose, dødsårsag, osv.) fra SST's registre bliver placeret hos SSI, således at de har ansvaret for indsamling af såvel relevante data samt biologiske prøver. Dette betyder, at SST ikke får noget valg i fht. udlevering af de registerdata, der er relevante for Biobankregisteret. Dermed vil Biobankregisteret opbevare alle relevante sundhedsdata, og SST vil ikke længere have kontrol over disse data.
2. Et andet kompromis er, at SST får ejerskabet over den nationale biobank og samtidig varetager driften af denne, således at sundhedsdata og biologiske prøver er samlet hos SST.
3. Et tredje kompromis indebærer en arbejdsdeling mellem SST og SSI, hvor SST beholder funktionen som registeransvarlig, samtidig med at SSI ligeledes har adgang til de relevante registerdata fra Biobankregisteret. Det vil sige at data kopieres fra SST's registre til SSI's Biobankregister.

Spørgsmål:

- Hvis du skulle vælge mellem et af de tre kompromiser, hvilket ville så være det mest funktionelle set ud fra SST's synspunkt?
- Hvorfor?

Vi mener det tredje kompromis vil være den bedste løsning for både SST og SSI, hvis den nationale biobank skal have en gavnvirkning, idet den har færrest konsekvenser for begge parter.

- Er det realistisk at det tredje kompromis kan udføres? (uddyb venligst hvorfor)
- Hvis du ikke kan forestille dig, at nogen af de ovenstående kompromiser kan realiseres, har du så andre forslag til hvorledes den nationale biobank kan etableres, således den har en gavnvirkning?
- Du fortalte tidligere, da vi ringede til dig, at kopiering af data ikke er hensigtsmæssig. Kan du uddybe hvori problemerne ved dette består?

Vi håber at du har mulighed for at besvare denne mail hurtigst muligt, da vi har en deadline for aflevering af vores speciale. På forhånd tak for din hjælp.

Med venlig hilsen Kate Z. Olesen og Ditte H. Kopp, 10. semester Sundhedsteknologi, Aalborg Universitet

Svar

Det er vigtigt at der genereres de data der skal genereres i forhold til brugerens behov, og at disse lagres på en sikker måde. Samtidig er det vigtigt, at dem der har brug for data og prøver rent faktisk har mulighed for at tilgå disse på en bureaukratisk, god, og let måde. Nøglen til dette er helt sikkert samarbejde mellem de indblandede parter, men det er vigtigt for mig at pointere, at der ikke er én instans, der kan håndtere den her biobank. Der er behov for både vores og SSI's ekspertressourcer, fordi vi har hver vores faglighed, hvorfor vi sammen kan finde en løsning på etableringen af den her biobank.

I er inde på noget af det rigtige med at lade alle instanser bidrage med hver deres faglighed, hvilket er helt central for at lave en hensigtsmæssig løsning for brugeren.

Kopiering vil medføre at data ligger flere steder, hvilket er uhensigtsmæssigt, og samtidig vil SSI have brug for konstant vejledning fra os, hvis de skal have vores data. Ligeledes skal SSI kunne stå inde for vores data, og er det nu også hensigtsmæssigt. Overordnet kan man sige, at vil man give forskere den her gode service, så må man udnytte alle instansers kompetencer.

G.4 John Møller-Jensen

Email

Kære John Møller-Jensen

Vi er to specialestuderende fra Aalborg Universitet, der studerer Medicinsk Informatik. Vi er af Lars Lilholt fra CSC blevet henvist til at kontakte dig, idet vi mangler hjælp i forbindelse med test af vores specialeprojekt.

Vores speciale omhandler datahåndteringen i den kommende nationale biobank i Danmark, der består af et biobankregister til datahåndtering samt en prøvebank til håndtering af biologisk prøver. I vores projekt fokuserer vi udelukkende på et konceptuelt design af biobankregisteret og har derfor ikke foretaget nogen implementering. Vi har derfor svært ved at validere de metoder vi anvender til at designe systemet, og har derfor været nødt til at "tænke ud af boksen" ang. test. Vi kunne godt tænke os at få valideret de standardmetoder vi benytter i design eller få et bud på, hvordan et sådan system ellers kan designes. I denne sammenhæng vil vi høre om du kan afse tid til at hjælpe os med dette?

Hvis du har tid og lyst, forestiller vi os at du kan danne dig et overblik over systemets formål og design ved skimme vores arbejdsblade for visionen for datahåndtering, systemets dele og design.

Vi håber at høre fra dig.

Med venlig hilsen Kate Olesen og Ditte Kopp

Svar

Efterfølgende tog John Møller-Jensen telefonisk kontakt, og følgende spørgsmål blev besvaret:

- Er de identificerede standardmetoder teknisk gangbare? Metoderne er som sådan ok, men de er meget pragmatiske.
- Kunne der med fordel have været anvendt andre standardmetoder? Der kunne anvendes en anden opdateringsmekanisme, der vil være mere effektiv og ligeledes dyrere. Herved ville der heller ikke blive tildelt et nyt ID for hver opdatering, hvilket er upraktisk da det gør det umuligt for forskere at koble nye og gamle data.
- Er der noget vi mangler at forholde os til med hensyn til data og standardmetoder? I burde lave en forbindelse til CPR-registeret, således at det er muligt at tjekke de indkomne CPR-numre med CPR-numre i CPR-registeret, idet der ofte sker fejl under indberetelse til PAS.