

## Udviklingsplan for CPR-systemet 2012 – 2015

### 1. Indledning

Det Centrale Personregister (CPR) indeholder grundlæggende personoplysninger om 9,2 mio. mennesker, som har tilknytning til det danske samfund. Siden CPR's oprettelse i 1968 har formålet med registret været, at registrets data skal stilles til rådighed for og offentlige myndigheder med en berettiget interesse i oplysningerne. I flere årtier er data også blevet videregivet til private virksomheder mv.

Data i CPR indberettes i dag online af de grundregistrerende myndigheder via CPR's online systemer, eller via CPR's web services, som er indbygget i myndighedernes sagsbehandlingssystemer. Det gælder bl.a. for statsforvaltningerne, Indfødsretskontoret og ikke mindst for Kirkeministeriets Personregistrering (PR), som er en integreret del af CPR-systemet.

Andre data indberettes i CPR af borgerne ved brug selvbetjeningsløsninger. Dette sker bl.a. på Personregistrering.dk og via kommunale flytteløsninger.

Anvendelsen af CPR's data, som generelt betragtes som værende af høj kvalitet, er i dag massiv i såvel den offentlige som den private sektor. CPR's dataleverancer har således i dag vital betydning for mange af landets kritiske it-systemer, og CPR spiller sammen med landets øvrige grundregistre en væsentlig rolle i udmøntningen af den fællesoffentlige digitaliseringsstrategi i de kommende år.

CPR-systemet hviler i dag på et robust, stabilt og velafprøvet teknisk grundlag, hvor de centrale teknologier, database og styresystemer er placeret på en mainframe server. CPR er karakteriseret ved stor driftsstabilitet og regelmæssighed i de daglige dataleverancer.

Denne driftsstabilitet skal opretholdes, men samtidig skal CPR-systemet følge med den teknologiske udvikling og til stadighed indrettes således, at omkostninger til drift, vedligehold og udvikling kan holdes på markedsniveau gennem konkurrenceudsættelse.

### 2. Grundlaget for udviklingsplanen

CPR-kontoret har gennem CPR-systemets levetid løbende vedligeholdt en flerårig udviklingsplan for systemet. Med henblik på at udarbejde en ny udviklingsplan for årene 2012-2015 har CPR-kontoret og Kirkeministeriet i efteråret 2011 i regi af CPR's udviklingsgruppe gennemført en analyse sammen med den nuværende driftsleverandør CSC Danmark A/S, som beskriver de mulige, konkrete veje og initiativer for udvikling og drift af CPR/PR. I forbindelse med analysen er CPR-kontoret og Kirkeministeriet bistået af Rambøll Management A/S.

Som udgangspunkt for analysen har CPR-kontoret og Kirkeministeriet overordnet set ønsket en ny model for CPR-systemet, som kan tilbyde følgende egenskaber:

- Der ønskes en platform og teknologi, som giver billigere drift og mere enkel vedligehold
- Der ønskes en platform og teknologi, som gør det lettere at videreudvikle applikationerne på sigt.
- Der ønskes en platform og en teknologi, som gør det lettere at konkurrenceudsætte drift, vedligeholdelse og videreudvikling af løsningen.

På grundlag af analysen har CPR-kontoret og Kirkeministeriet besluttet sig for en strategi, som er kendetegnet ved:

- Begrænsning af risiko ved at opdele udviklingsstrategien i et antal ikke for store projekter
- Sikring af at muligheden for konkurrenceudsættelse forbedres væsentligt
- Trinvis migrering til den nye driftsplatform for derved at sikre lav risiko
- Bibeholdelse af Natural som udviklingsværktøj i en periode
- Etablering af en ny domæne model og database for at sikre optimal fleksibilitet ved fremtidig udvikling
- Trinvis modernisering af kritiske systemer
- Mulighed for løbende at tilpasse aktivitetsplan efter udefrakommende krav som f.eks. en fællesoffentlig datafordeler eller en fællesoffentlig domænemodel, jf. regeringens digitaliseringsstrategi

CPR-kontoret og Kirkeministeriet har endvidere besluttet sig for at implementere strategien på en sådan måde, at der ikke forudsættes store spidsbelastninger - hverken hos CPR-kontoret, Kirkeministeriet eller CSC, ligesom de fleste aktiviteter indledes med en et såkaldt proof of concept (POC), hvor løsningen over en periode på 2-3 måneder beskrives, og hvor kritiske elementer afprøves i praksis. POC verificerer de hypoteser, der er opstillet i udviklingsstrategien og giver grundlag for at udfærdige en udviklingsaftale for projektet.

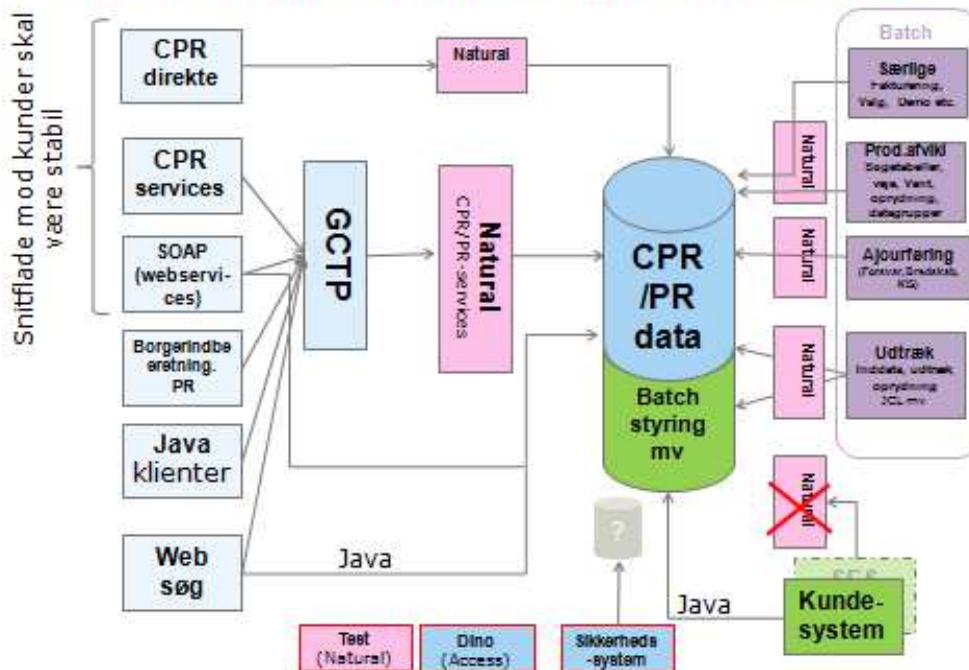
### **3. Det eksisterende løsningslandskab**

CPR-systemet skal give brugere over hele landet adgang til at registrere, søge og få leveret oplysninger fra CPR. CPR-systemet indeholder aktuelle og historiske personoplysninger om hele den danske befolkning samt oplysninger om samtlige veje, kommuner og myndigheder i Danmark.

CPR-systemets formål er dels at opbevare disse data, dels at vedligeholde dem og videregive dem på en teknisk og økonomisk forsvarlig måde.

CPR-databasen er centralt placeret og tilhører CPR-kontoret og Kirkeministeriet. CPR-systemet fungerer som forbindelsesled og leverandør i forhold til betjening af offentlige og private kunder, samordningen mellem stat og kommuner, valg, planlægning, statistik og systemeksport. I samarbejde med Kirkeministeriet er Personregistreringssystemet (indtil april 2007 kaldt DNK - Den Ny Kirkebog) en integreret del af drift og forvaltning af CPR-systemet, mens Kirkeministeriet alene står for videreudvikling af Personregistreringssystemet (herefter forkortet PR).

## SYSTEMOVERSIGT-INTEGRATIONSPUNKTER



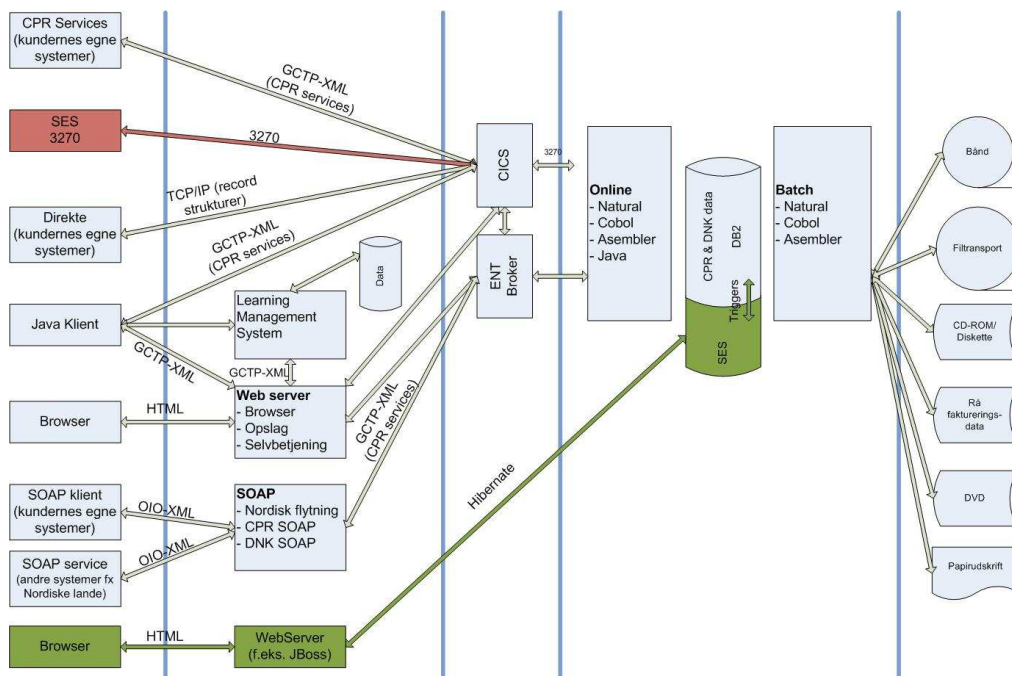
**Figur 1 - Overordnet beskrivelse af CPR/DNK systemet og dets anvendelsesområder**

Figuren ovenfor viser de overordnede forretningsområder, som systemet dækker, herunder hvilke aktører disse forretningsområder understøtter og integrerer imod.

Det bemærkes, at det såkaldte GCTP lag danner et integrationslag mellem de enkelte komponenter/applikationer og de bagvedliggende systemer og øvrige komponenter. Eneste undtagelse er CPR Direkte og de simple søgninger via Websøg, som ikke kører via GCTP. Denne overordnede struktur giver mulighed for en struktureret og trinvis omlægning af systemerne, da de ikke er vævet ind i hinanden, men logiske adskilte.

### 3.1. Tilføjelsen af Kundesystemet

Ud over det ovenfor beskrevne systemlandskab er et nyt Kundesystem p.t. under udvikling. Kundesystemet er CPR-kontorets værktøj til at administrere dataleverancer og fakturering. Det nye Kundesystem, som afløser et 20 år gammelt tegnbaseret system, er baseret på et nyt rammesystem, et nyt udviklingsprincip og udfordrer den nuværende databasemodel.



**Figur 2 - Kundesystemets indplacering i CPR arkitekturen**

Figuren ovenfor viser, hvorledes Kundesystemet bygges som en del af CPR-systemet, således at det sikres, at Kundesystemet på den ene side integrerer til det eksisterende, men på den anden side også med beskedne ændringer kan anvendes med en forandret datamodel.

Kundesystemet er en web applikation, opbygget i Grails oven på en Java baseret web server. Der er etableret en ny datamodel og nye tabeller i den nuværende DB2 database for Kundesystemets data. Dette betyder, at der enkelt kan ske en tilpasning, når der implementeres en ny fælles datamodel. Der vil blive benyttet triggers til at synkronisere de nuværende tabeller med relevant indhold i de nye tabeller.

### 3.2. Rambøll Managements observationer

I forbindelse med arbejdet i Udviklingsgruppen har Rambøll Management på baggrund af gennemgangen af de forskellige systemer og hjælpeværktøjer foretaget nedenstående væsentlige observationer:

Det nuværende CPR-system stammer fra 60'erne, men har flere gange været gennemgribende moderniseret, f. eks. i begyndelsen af 90'erne og 00'erne.

Systemet er præget af, at det skal rumme historik vedrørende registreringer tilbage til starten.

Systemets grundlæggende arkitektur bygger på en opdeling i komponenter med veldefinerede grænseflader og genbrug af funktionalitet. Det giver mulighed for udvikling af systemet komponent for komponent.

Systemet er karakteriseret ved forholdsvis få daglige opdateringer (12.000) og ved at udtræk til kundernes systemer omfatter 2-3 mio. personer dagligt med 15-20 mio. personer ved månedsskifte. Der er høje krav til datakvaliteten. Kravene til tidstro registrering er ikke p.t. på samme niveau som i bankernes systemer. Den samlede mængde data er begrænset set i forhold til nutidens teknologi. Det betyder, at systemet ikke længere stiller krav, som kun kan leveres af en mainframe.

Omkring CPR-systemet er der bygget en række hjælpesystemer såsom DINO (dokumentation og konfigurationsstyring), en egenudviklet teststyringsmodel, et egenudviklet kundestyringsystem, et egenudviklet system til udtræk af og dannelse af faktureringsgrundlag, der leveres som lister til CSCs SAP, og en egenudviklet sikkerhedsmodel. Disse hjælpesystemer bevirker, at CPR systemet er godt dokumenteret til brug i drift og udvikling samt godt administreret.

Det samlede system har dog også nogle svagheder, som bør håndteres i den fremtidige udvikling.

Som følge af systemets lange historie er der træk, som skyldes hensyn til de tekniske muligheder, da den aktuelle arkitektur og konkrete komponenter blev designet og udviklet. I forhold til mulighederne i dagens teknologi, har CPR-systemet følgende svagheder:

- Håndtering af historik (tidspunkt for opdatering og for gyldighed mv) er designet med henblik på at spare lagerplads.
- Håndtering af ændringsudtræk (der oprettes tabeller med dagens ændringer) er designet for at spare cpu forbrug.
- Håndtering af søgninger er håndteret med hjælpetabeller for at håndtere søgninger efter forskellige stavemåder og håndtere forskelle i stavemåder over tid
- Håndtering af opdateringer, som hindrer at udtræk kan foregå parallelt med udtræk fra systemet.
- Manglende ensartethed i håndtering af ændringer.
- Manglende mulighed for at finde ændringer siden et givent tidspunkt. I dag hentes ændringerne dagligt. Det bevirker, at hvis en kunde ønsker ændringsudtræk dannes der et dagligt udtræk, uafhængigt af om kunden kun ønsker leverancen med en anden frekvens.
- Manglende mulighed for at gentage tidligere produktionskørsler. Tilføjelse af denne mulighed vil sikre bedre testmuligheder
- CPR-systemets kompleksitet betyder, at ændringer i krav til data (f. eks. tilføjelse af et ekstra felt) kræver en stor ressourceindsats til ændringer i selve databasen, til ændringer i ajourføringsdelen (skærm billeder, snitflader) og til udtræksdelen
- Et meget stort antal tabeller med få dataelementer i hver tabel, hvilket var fornuftigt da databasen blev designet. Antallet af tabeller vil blive reduceret kraftigt ved et redesign.
- Begrænset brug af indeks, hvilket også var fornuftigt da databasen blev designet

Dette betyder et behov for modernisering og forenkling af datamodellen samt de komponenter, der håndterer udtræk og søgninger.

Til test primært af de mange Natural-programmer er der som nævnt udviklet en teststyringsmodel, der også indebærer, at databasen indeholder mange testtabeller. Testmodellen indeholder mere end 3.000 Natural moduler og 235 tabeller. Når der er tabelændringer, skal der derfor rettes i mange Natural programmer. Der er behov for at udvikle en ny testmodel, hvor dog de mange testscenarier, der er indbygget i testmodellen, kan genbruges i et nyt værktøj og ved manuel test.

Dino benyttes i dag til opbevaring af dokumentation, til konfigurationsstyring af Natural programmer og som krydsreferencesystem. Systemet er baseret på en MS Access 2003 database. Systemet skal erstattes.

Der er ikke en ensartet metode til dokumentation af det samlede system, hvilket betyder, at mens dele af systemet har let tilgængelig dokumentation, har andre dele behov for forbedringer, ligesom der bør være en fælles dokumentationsmodel for det samlede system.

Den eksisterende Web baserede løsning til simple søgninger afvikles delvist på mainframe og delvist på Linux. Løsningen bruger imidlertid meget mainframe kapacitet og performance er ikke tilfredsstillende.

Batch-afviklingen, som løser flere opgaver ud over udtræk, er præget af de tekniske bindinger til mainframen og af, at batchafviklingen har en lang forhistorie. Der bør igangsættes en oprydning af

funktioner og moduler, der ikke længere anvendes, så disse ikke skal migreres til ny platform. Herefter skal dokumentationen opdateres.

Den stigende anvendelse af Citrix løsninger hos CPR brugerne giver større og større problemer, da manuel indgriben er nødvendig fra administratorer, hver gang der distribueres en ny Java klient.

Etablering af de forskellige SOAP-services og webbaserede selvbetjeningsløsninger er sket over flere år, og der er valgt forskellige arkitekturer som følge af forskellige tekniske hensyn og behov. Ved nyudvikling på området bør der overvejes en konsolidering af løsningerne på dette område. Det skal indgå i udviklingsstrategien at sikre, "at drift, vedligehold og udvikling reelt kan konkurrenceudsættes, ligesom en fremtidig løsning bør åbne mulighed for at delelementer eller nye moduler kan leveres fra forskellig side". Ud over de ovennævnte aktiviteter skal der findes en løsning på, at CSC anvender delkomponenter, som ikke uden videre kan overtages af en eventuel ny leverandør efter et udbud.

#### **4. Udviklingsplanens hovedpunkter**

##### **Migrering af Natural til midrange**

CPR-systemet afvikles i dag primært på en mainframe server, som af mange betegnes som en forholdsvis dyr platform, og hvor konkurrencen for drift af denne platform i Danmark må betegnes som begrænset. Som anført af Ramøll er der ikke forhold omkring CPR-systemet, som tilsiger, at systemet nødvendigvis skal afvikles på en mainframe platform.

CPR-kontoret og Kirkeministeriet har derfor sammen med den nuværende driftsleverandør, CSC, etableret et testmiljø og gennemført en prøvemigrering af CPR-systemets online applikationer og DB2 database til en midrange platform baseret på et antal HP servere med Linux som operativsystem.

Resultatet af prøvemigreringen har vist, at systemet skal undergå en række tilpasning i forbindelse med en migrering, men at disse tilpasninger ikke er af uoverstigelig karakter.

CPR-kontoret og Kirkeministeriet har på dette grundlag besluttet at fortsætte arbejdet med at klargøre en migrering af CPR-systemets on line applikationer til en Linux platform. Skiftet af platform forventes at medføre mærkbare besparelser på driften, bl.a. som følge af den mere fordelagtige konkurrencesituation på midrange markedet.

##### **Oprydning i batch**

CPR-systemets batch har det primære formål at afvikle en stribe programmer, som fremhenter, sorterer og fordeler data til CPR's kunder. På daglig basis afvikles over 400 ændringsudtræk, som har til formål at fremskaffe de personer, som har fået ændret deres data i løbet af dagen.

Systemet er paralleliseret, så den daglige afvikling er nede på ca. 50 minutter, hvorunder der fremskaffes og fordeles ca. 350.000 personer til de 400 modtagere. I alt leveres der ca. 1.7 mio. records i den daglige kørsel.

Batchflowet sikrer endvidere, at datagrundlaget til fakturering, dannelsen af benyttelsesstatistikker og opdatering af sikkerhedslogs bliver dannet.

Det kørende batch flow omfatter over 1.300 aktive batchjobs, som for den største dels vedkommende er kodet i Natural.

CPR-kontoret og Kirkeministeriet har sammen med den nuværende driftsleverandør, CSC, foretaget en overordnet gennemgang af batchjobs og batchprogrammer i produktionsmiljøet – primært fokus har været de daglige jobs. Resultatet af gennemgangen har vist, at batchprogrammerne skal undergå en række tilpasninger i forbindelse med en migrering, men at disse tilpasninger ikke er af uoverstigelig karakter.

CPR-kontoret og Kirkeministeriet har på dette grundlag besluttet at fortsætte arbejdet med at klargøre en migrering af CPR-systemets batchprogrammer til en Linux platform og gennemfører i den forbindelse en oprydning i programmerne som anført af Rambøll.

### **Adgangsstyring**

Det bestående adgangsstyringssystem er meget svært at anvende, og der er stor personafhængighed til en nøgleperson i CPR-kontoret.

CPR-kontoret har gennemført en selvstændig foranalyse vedrørende etablering af et nyt adgangsstyringssystem. På baggrund af analysen er det besluttet, at der udarbejdes en løsningsbeskrivelse samtidig med, at der i Grails udvikles en prototype på brugerdialogen i det nye system.

Det nye adgangsstyringssystem bygger på den i foranalysen definerede sikkerhedsmodel. Det er i foranalysen vurderet at en standard LDAP sikkerhedsmodel ikke giver den funktionalitet som CPR-kontoret har brug for.

Efter implementering forventes RACF i en periode fortsat benyttet til kontrol af log-on. Senere omlægges denne kontrol formentligt til et standard identity management system på midrange.

### **Ny domæne model/database**

CPR's nuværende datamodel blev implementeret i 1991 ved omlægning fra IMS til DB2. Der er siden indarbejdet ændringer i modellen – men modellen er grundlæggende den samme blot udvidet.

Ved tilføjelse af ny funktionalitet er der ofte udarbejdet nyt datagrundlag til understøttelse af funktionaliteten. Dette har medført at den eksisterende databasemodel indeholder en del redundante transaktionsdata. Manglen på en generel CPR model gør nyudvikling og rettelser i det eksisterende system både besværligt og dyrt samt indebærer øget risiko for fejl.

Der er følgende øvrige forretningsmæssige bevæggrunde for, at udarbejdelse af en ny datamodel indgår i CPR-systemets udviklingsplan:

- Referentiel Integrity

Referentiel Integrity(RI) er af historiske grunde ikke implementeret i nuværende datamodel. Ved implementering af RI i en datamodel sikres, at relationer mellem data er sammenhængende – dvs. det højner datakvaliteten. Ved implementering af RI er det ikke applikationen, der skal sikre disse sammenhænge – men databasestyresystemet. RI kan ikke implementeres i eksisterende database uden væsentlige applikationsændringer.

- Dataformater

Felter, der indeholder datoer, skal i en ny datamodel være defineret som date format. I dag er der en del start- og slutdatoer, der er defineret som numerisk 12, idet de foruden dato indeholder klokkeslæt. Visse af datoerne kan indeholde værdien 00 i dag og/eller måned. Numeriske felter er

defineret som decimal-felter – dvs. et format, hvori der kan angives decimaler, hvilket ikke er relevant i forbindelse med data i CPR. Disse data skal defineres med korrekt format.

- Adresser

I CPR registreres en persons danske adresse, hvis personen har bopæl i Danmark. Adressen registreres med personens kommunekode, vejkode, husnummer osv. Har personen ikke dansk adresse, men skal registreres i CPR af andre årsager – fx skattemæssige forhold (sømandsskat), tildeles en fiktiv adresse til personen indeholdende en fiktiv kommunekode. Ved ny datamodel udgår disse "fiktive" adresser.

- Relationen mellem forældre/barn

En persons forældre registreres i dag i CPR på samme forekomst, dvs. en forekomst indeholdende Barn, Mor-oplysninger, Far-oplysninger. Ved en ny datamodel skal forældre være indeholdt i hver sin entitet.

- Registrering af ægteskab

Ved ajourføring af en vielse registreres der en forekomst indeholdende henvisning fra mand til kvinde, samt en anden forekomst med henvisning fra kvinde til mand. På denne måde registreres der gensidig henvisning, hvilket er redundante data. Ved design af ny datamodel fjernes gensidig henvisning.

- Tilpasning til moderne udviklingsværktøjer

Den eksisterende model gør det vanskeligt/umuligt at arbejde med moderne udviklingsværktøjer som f.eks. Grails, idet Grails baserer sig på RI/datasammenhænge mellem de forskellige tabeller/objekter. Den eksisterende database er desuden fremkommet i en tid, hvor struktureret programmering var den fremherskende udviklingsmetode. Udviklingsmetoder er i dag primært baseret på en objektorienteret tilgang. En database bør i dag afspejle dette for at kunne drage fuld fordel af de værktøjer, der er til rådighed. Moderne udviklingsværktøjer arbejder med datamodeller i form af domænemodeller, som afspejler de forretningsmæssige sammenhænge. En domænemodel er en integreret del af udviklingsprocessen hvorved at modellen og tilhørende database naturligt kommer til at udvikle sig i takt med nye krav/ændringer til CPR.

## 5. Udviklings- og supportværktøjer

Dette kapitel beskriver de primære udviklingsværktøjer, som foreslås anvendt i forbindelse med udviklingsstrategien. Det skal bemærkes at den reelle levetid for et udviklingsværktøj i dag typisk vil være under 10 år. Det er derfor væsentligt at vælge produkter der vurderes at kunne "porteres" samt løbende at opgradere værktøjer og systemer så de er ajour med markedet.

### Natural

Natural er i dag et nicheprodukt og markedsandelen for produktet reduceres stadig. Der er dog intet akut behov for at migrere værk fra Natural, da produkterne stadig vedligeholdes og supportes af leverandøren og da der stadig er en del store kunder, bl.a. i den finansielle sektor, også i Danmark.

Det væsentligste problem er at der i forhold til konkurrenceudsættelse kun er få udviklingsleverandører, der har ekspertise i udvikling og vedligeholdelse af systemer baseret på Natural.



## Enterprise Architect

Enterprise Architect er et værktøj, der benyttes til at modellere og designe IT systemer, og som understøtter hele produktudviklingens livscyklus; lige fra de overordnede forretningsmodeller over logiske datamodeller til fysiske ER-diagrammer.

Enterprise Architect er et visuelt værktøj, som understøtter model-drevet udvikling, hvor software bygges op med visuelle logiske modeller, hvorfra der automatisk kan genereres kørende kode. Dette giver en høj grad af fleksibilitet, da modellen udarbejdes uafhængigt af den underliggende platform. Det hjælper også i høj grad til at holde forretningen og den tekniske implementering adskilt.

Enterprise Architect understøtter flere forskellige arbejdsprocesser og indeholder følgende funktioner.

- Er baseret på UML notifikationen
- Understøtter modellering og design i teams.
- Understøtter den fulde life cycle
  - Forretnings- og procesmodellering
  - System design
  - Enterprise arkitektur
  - Krav håndtering og mapning
  - Software design
  - Kode generering
  - Test
  - Dokumentation

Enterprise Architect er således med til at sikre, at der er sammenhæng, hele vejen fra designet af de forretningsprocesser man ønsker understøttet, over den kode som implementerer funktionaliteten, til test af det endelige system.

Det betyder også, at dokumentationen bliver mere korrekt og sammenhængen til det implementerede klarere.

Enterprise Architect giver mulighed for at eksportere modeller til RTF og HTML som dokumentation og understøtter mange officielle standard modellerings-specifikationer som UML, SysML, DDS og BPMN.

Enterprise Architect benyttes bl.a. på Kundesystemet.

## Spring Framework

Spring fra firmaet SpringSource (nu en del af VMware) er et Open Source framework, som har eksisteret siden 2003. Det er skabt til at simplificere udvikling af Enterprise Java applikationer, så de ikke er bundet til en bestemt applikationsserver fra en bestemt leverandør. Anvendelse af Spring øger produktiviteten og skaber mere fleksibilitet og større grad af uafhængighed. Spring er en af grundstenene i web frameworket Grails, som benyttes til at implementere kundesystemet.

## Jira

Jira er et meget udbredt projektstyrings- og fejlhåndteringssystem, som gør det muligt at styre og overvåge alle dele af moderne udviklingsprojekter. Jira benyttes til at logge fejl, udestående opgaver, ressourcer osv. I forbindelse med udviklingsprojekter. Jira kan f.eks. linke direkte mellem bugs og kode, så fejl og rettelser kan relateres til builds, versionering og kodelinjer.

Derudover indeholder Jira også stærke værktøjer til at udtrække statistik og rapporter, så det også er muligt at styre og overvåge håndteringen og løsningen af issues og fejl i projektet og produktet. Jira gør udviklingsprocessen synlig både for CSC som leverandør og for Kunden.

Jira benyttes allerede i forbindelse med udviklingen af CPR's nye kundesystem.