



■ ■ ■ ■ En rapport for
**Forsknings- og
Innovationsstyrelsen**

En national forskningsdatabase

21-12 2007

Engagement: 221998430

Table of Contents

1.0	Baggrund	4
2.0	Metode.....	5
3.0	Systemlandskabet i dag	6
3.1	PURE.....	6
3.2	Orbit.....	7
3.3	Den Danske Forskningsdatabase (DDF).....	8
3.4	Samlet vurdering af det eksisterende systemlandskab	9
4.0	Internationale erfaringer.....	10
4.1	Norge.....	10
4.2	England	11
4.3	Holland	12
4.4	Opsamling af internationale erfaringer	12
5.0	Rammer for en dansk forskningsdatabase.....	13
5.1	Rammer udstukket af det faglige udvalg	13
5.2	Tidsmæssige bindinger for etableringen af en fælles forskningsdatabase.....	14
6.0	Fremtidige modeller for en national forskningsdatabase	15
6.1	Scenarie 1: fuld centralisering	16
6.2	Scenarie 2: Decentral model	17
7.0	Validering og etablering af hjælpedatabaser	20
7.1	Decentral og central validering	22
8.0	Aktiviteter på kort sigt.....	25
9.0	Opgaver og organisering af en central organisation.....	26
10.0	Krav til en centraliseret løsning	29
10.1	Arkitekturtegning til system.....	29
10.2	Funktionelle moduler:	29



■ ■ ■ ■ **Sammenfatning**

Sammenfatning

Regeringen har besluttet at etablere en bibliometrisk forskningsindikator, som skal anvendes i forbindelse med tildeling af bevillinger til universiteterne i Danmark. Etableringen af forskningsindikatoren er et politisk følsomt projekt og udover udformningen af selve indikatoren bliver det kritisk for indikatorens succes, at man kan etablere et troværdigt datagrundlag for den.

I regi af universiteterne og DEFF findes der i dag systemer, som til sammen kan anvendes til at producere den bibliometriske forskningsindikator. Gartner anbefaler dog en række justeringer til systemerne, for at give mulighed for at forbedre datakvaliteten. Det drejer sig om etableringen af en hjælpedatabase med den autoritative liste over publikationskanaler, samt en hjælpedatabase, som indeholder et udtræk fra ISI eller Scopus. For at disse hjælpedatabaser fungerer, kræver det desuden små tilpasninger af PURE og Orbit.

For at kunne etablere den bibliometriske forskningsindikator for publikationer i 2008, bør disse initiativer sættes i gang, uanset hvilken model man vælger på længere sigt.

Etableringen af hjælpedatabaserne bør kunne gennemføres for under 500.000,-, og dermed uden at kræve et udbud og tilpasningerne til PURE og Orbit bør kunne gennemføres for under 50.000,- pr. institution. Med denne investering er det muligt at etablere et IT-mæssigt grundlag for en bibliometrisk forskningsindikator. For at nedbringe omkostningerne til valideringsprocesser bør man dog på sigt foretage yderligere investeringer.

Det er Gartners vurdering, at der er behov for at etablere en central organisation f.eks. i regi af F&I eller DEFF, som kan varetage en række opgaver bl.a. knyttet til vedligeholdelse af den centrale database og hjælpedatabaser, validering og deduplicering af registreringer, samt udregning af indikatoren.

Man kan i Danmark fremadrettet vælge ét af to scenarier for systemmæssigt at understøtte registreringer af dansk forskning:

1. Erstat det nuværende systemkompleks med ét centralt system, f.eks. i regi af UNI-IT og gennemfør udbud på etablering og drift af et sådant system.
2. Bibehold det eksisterende systemkompleks og gennemfør et udbud af driften af den nationale forskningsdatabase, baseret på beta-versionen af det system som i dag kendes under navnet DDF.

Det første scenarie vurderes samlet set at lede til besparelser på IT-drift og valideringsprocesser, men kræver etableringen af en styringsmodel og en investering i at gennemføre et udbud.

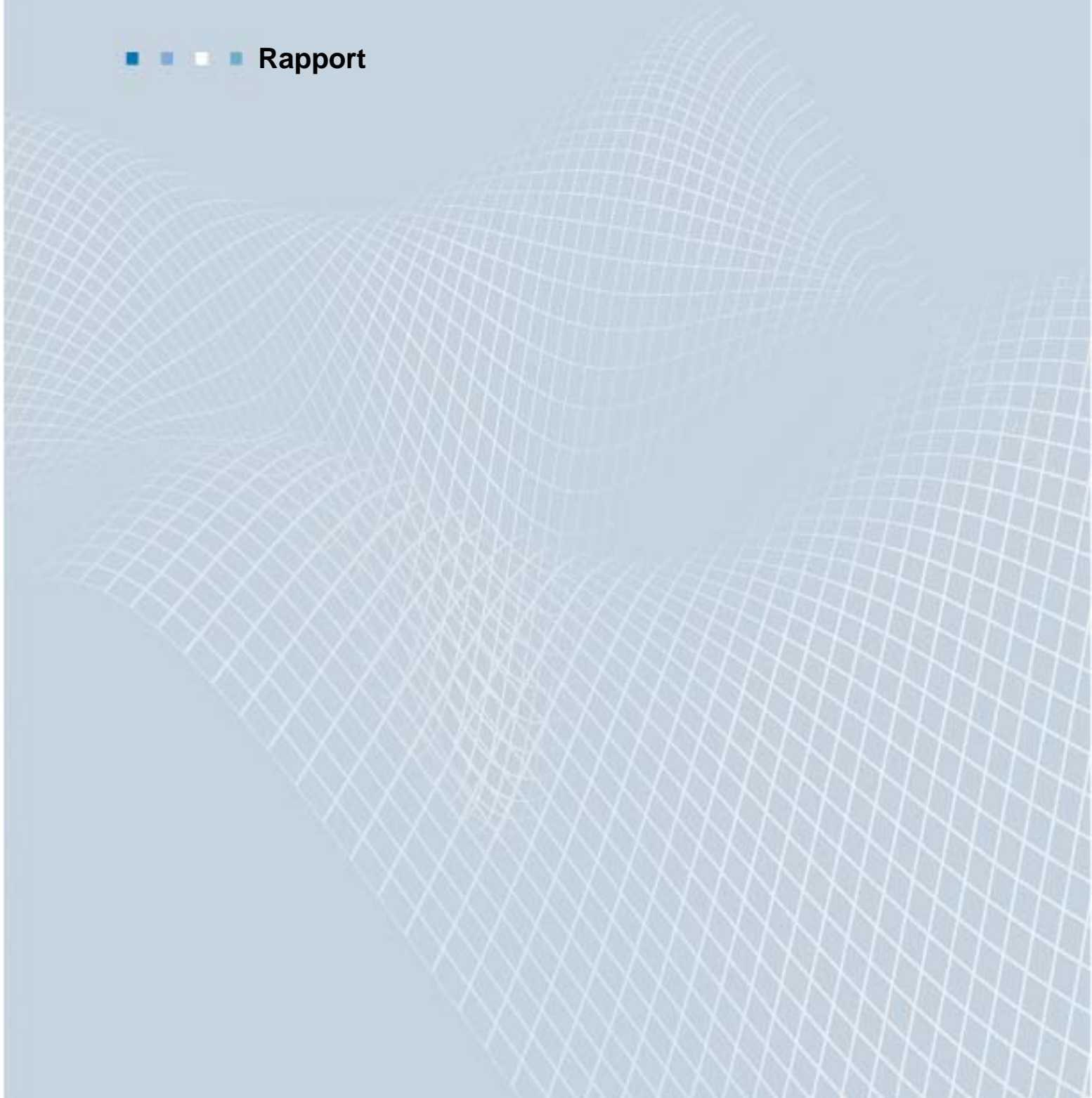
Gartner vurderer at en central løsning vil kunne etableres for ca. 2 mio. kr. og at IT-systemet vil kunne driftes for ca. 400.000,- pr. år.

En central løsning vil betyde en række afledte investeringsomkostninger hos nogle af universiteterne til tilpasning af formidlingssystemer, som i dag er integreret med PURE og Orbit. Det afledte investeringsbehov vurderes at kunne udgøre ca. 500.000,-

Det decentrale scenarie vil på sigt kunne løse opgaven som den gør nu, samt understøtte den bibliometriske forskningsindikator med en vurderet investering på ca. 500.000,- og minimal risiko, men med en samlet driftsomkostning på årligt ca. 1,7 mio. kr..

I forbindelse med udbuddet på driften af den nationale forskningsdatabase i det decentrale scenarie, bør man forvente et årligt driftsvederlag på ca. 100.000,-. Hertil kommer omkostninger til forvaltning og løbende videreudvikling.

■ ■ ■ ■ Rapport



1.0 Baggrund

Som konsekvens af regeringens Globaliseringsstrategi har Videnskabsministeriet besluttet at udvikle en bibliometrisk forskningsindikator, der skal identificere og måle forskningspublicering på tværs af institutioner.

Den bibliometriske forskningsindikator skal bl.a. anvendes i forbindelse med fastlæggelse af bevillinger til universiteterne.

I den forbindelse er der nedsat et teknisk udvalg, som skal forberede etableringen af et IT-system som kan understøtte forskningsindikatoren. Det tekniske udvalg har bedt Gartner foretage en undersøgelse, som skal forberede etableringen af et sådant system. I kommissoriet for undersøgelsen står bl.a.:

Den eksterne konsulentundersøgelse har til formål at undersøge samt udarbejde et forslag til udvikling af en database og et system til indsamling, registrering og validering af publiceringsdata

Konsulentundersøgelsen skal fokusere på en række forskellige delundersøgelser, som vurderes at være særlig relevante. Det drejer sig om følgende:

- *Struktur, organisering og drift af den danske forskningsdatabase.*
- *Etablering af hjælpedatabaser og autoritetsdata.*
- *Validering.*
- *Deduplicering.*

2.0 Metode

Gartner har gennemført undersøgelsen på baggrund af bl.a. følgende:

- Materiale stillet til rådighed af det tekniske udvalg, herunder det tekniske udvalgs notat: "Teknisk udvalgs grundpapir".
- Assistance fra Adrian Price, KU, som bl.a. har indsamlet data vedrørende de eksisterende systemer til registrering af forskning.
- Gartners analytikere indenfor higher education.

Desuden er der gennemført interviews med systemejere af de tre eksisterende systemer, som det er relevant at undersøge i denne forbindelse:

- Claus Vesterager, RUC, formand for brugergruppen for PURE.
- Mogens Sandfær, DTV, vedrørende Orbit og DDF(Den Danske Forskningsdatabase)
- Alfred Heller, DTV vedrørende Orbit
- Mikael Elbæk, DTV vedrørende DDF (Den Danske Forskningsdatabase).

Gartner har ved to lejligheder præsenteret og diskuteret foreløbige observationer og konklusioner med teknisk udvalg med det formål at validere dem.

3.0 Systemlandskabet i dag

Den danske universitetsverden anvender to forskellige systemer til registrering af forskningsaktivitet, PURE og Orbit.

Orbit anvendes på Danmarks Tekniske Universitet og på Forsvarsakademiet, mens PURE anvendes på de øvrige universiteter.

3.1 PURE

PURE anvendes til at registrere forskningspublikationer samt på nogle universiteter til forskningsaktiviteter og forskningsprojekter. PURE anvendes både til formidling og til udarbejdelse af statistikker over forsknings- og publiceringsaktivitet. Data fra PURE kan således genanvendes i universiteternes andre Internet systemer, sådan at den enkelte forsker eller den enkelte institution eksempelvis relativt let kan integrere publikationsoplysninger fra PURE i elektroniske CV'er eller andre formidlingssammenhænge.

Leverandørforhold

PURE Projektet er oprindeligt vokset ud af et projekt på Aalborg universitet. Det blev til et udviklingsprojekt med deltagelse af 4 universiteter (KVL, HH i Århus, AAU, RUC). Projektsamarbejdet mellem de fire universiteter var støttet af DEFF.

Udviklingen af PURE forestås af firmaet Atira A/S, der supporterer og løbende udvikler og implementerer nye elementer til systemet. Atira A/S er et lille dansk systemudviklingshus, som markedsfører PURE til lærestalter internationalt.

Det overordnede ansvar for udviklingen af systemet ligger i en styregruppe, som har repræsentanter fra alle de involverede universiteter og hvor Atira A/S har observatørstatus. Udviklingsomkostninger fordeles mellem universiteterne efter samme model som licensen, og fastsættes ud fra antallet af forskningsårsværk på universiteterne. Det vil sige, at de største universiteter betaler mest, men også har flest "stemmer" i samarbejdet.

Finansieringen af systemudviklingen er fleksibel, sådan at hvis et enkelt universitet har ønsker om specialudviklede løsninger, så kan det pågældende universitet selv betale for dem, såfremt den ønskede løsning ikke griber forstyrrende ind i systemkomponenter, der anvendes af de øvrige universiteter.

Nyudviklinger besluttet i styregruppen og dimensioneres og budgetteres i et samarbejde mellem et teknisk udvalg og Atira. Det tekniske udvalg har ligesom styregruppen medlemmer fra alle universiteter. Modellen betyder, at det fra gang til gang forhandles på plads, hvorvidt ændringer er en del af licensforholdet, eller er specialudviklinger.

Systemet købes som en grundpakke, hvortil der kan købes en række moduler i tilvalg. Det er bl.a. et Rapport modul, Bibliometri modul, Studenter-projekt modul og Presseklip modul.

Driftsmodel

PURE drives forskelligt på universiteterne. På både Århus og København findes der, bl.a. som følge af universitetsfusionerne, flere parallelle installationer. Der arbejdes på at samle installationerne til en enkelt pr. universitet. På RUC og AAU drives PURE af biblioteket, som har sine egne servere, der drives uafhængigt af universitetets IT-afdeling.

Workflow og datakvalitet

PURE anvendes på alle universiteter, undtaget DTV og ITU, og inddatering af poster og anvendelse af f.eks. statistik over publikationsdata sker med stor variation. Inden for de forskellige universiteter er der forskelle, helt ned på enkelt-institut-niveau.

Generelt er det enten den enkelte forsker eller en sekretær på instituttet, som forestår inddateringen i systemet. Dernæst godkendes inddateringerne, typisk af institutlederen.

På CBS er det biblioteket, der forestår inddateringen, med udgangspunkt i fysiske eksemplarer af publikationerne. I disse sammenhænge har forskerne således ingen andel i registreringsprocessen.

Af formelle grunde er det dekanen, der foretager den endelige, faglige validering, dvs. afgør om publikationskanaler kan antages som videnskabelige.

Den bibliografiske validering håndteres en del steder af en sekretær, men mange steder er det lagt ud til bibliotekarer på institutbiblioteker eller (på RUC) forskningsbiblioteket.

Der findes ligeledes forskellige modeller for, i hvilken grad resultaterne anvendes i ledelsen af de enkelte institutter. Generelt anvendes publikationsdata som en fast dimension i ledelsen af institutterne. Nogle steder findes der desuden faste rutiner, som foreskriver regelmæssige indberetninger, f.eks. for at muliggøre kvartalsvise opgørelser for større enheder, f.eks. på fakultetsniveau. Andre steder opereres der med en årlig deadline.

Teknisk arkitektur

PURE er udviklet i Java, på en open source platform og det er Gartners overordnede vurdering, at systemet har en sund arkitektur og er bygget på tidssvarende teknologier, som man kan basere en videreudvikling på.

3.2 Orbit

Orbit bruges til at registrere forskningsaktiviteter (bl.a. forskningsprojekter) samt forskningspublikationer. Der er lagt vægt på at data fra Orbit kan genanvendes i DTV's andre internet systemer, sådan at den enkelte forsker eller den enkelte institution eksempelvis relativt let kan integrere publikations- og projektoplysninger fra Orbit i elektroniske CV'er eller andre formidlingspublikationer.

Systemet er integreret med det eksisterende projekt- og personaleadministrationssystem, Scanpas.

Leverandørforhold

Orbit udvikles af ePub-enheden på DTV.

Orbit er integreret i DTV's langsigtede planer om udviklingen af et forskningsarkiveringssystem, der skal omfatte både publikationsdata, oplysninger om forskningsprojekter, elektroniske kopier af forskningspublikationer samt rådatasæt fra forskningsprojekter (e.g. målingsdata).

Driftsmodel

Orbit drives på DTV af enheden ePub, som oprindeligt har udviklet systemet, og som forsat udvikler på det.

Workflow og datakvalitet

Orbit giver mulighed for flere forskellige workflows. Den mest almindelige er at indtastningen i Orbit foretages af den enkelte forsker selv. Dernæst kontrolleres posten på instituttet, inden

den valideres endeligt af en bibliotekar. Begge disse valideringer er manuelle. Det er først med den endelige validering at posten er indført i Orbit og frigives til f.eks. høstning af DDF.

Der hentes data fra ISI, baseret på det enkelte instituts publikationsprofil. Det betyder, at der forefindes en lokal tidsskriftsfortegnelse med de mest benyttede tidsskrifter på instituttet, som stilles til rådighed i forbindelse med indtastninger i systemet. Systemet bidrager bl.a. til ensartede tidsskrifts- og forlagsoplysninger. Mellem 30 og 50 % af publikationerne i Orbit valideres via ISI.

Skønt poster gennemgår et relativt detaljeret check på flere niveauer, skønnes det at mellem 1 og 3 procent af alle poster indeholder fejl.

Teknisk arkitektur

Orbit er ligesom PURE baseret på Java på en open source platform, og det er Gartners overordnede vurdering, at systemet har en sund og fleksibel arkitektur.

3.3 Den Danske Forskningsdatabase (DDF)

Den Danske Forskningsdatabase har eksisteret som løsning under forskellige navne med forskellige teknologier siden 1988. DDF har fra starten været tænkt som en portal, hvorfra man samlet kunne få adgang til oplysninger om dansk forskning. DDF gennemgik omkring 2002 et udviklingsprojekt, som bl.a. betød at systemet fik en række forbedrede søgefaciliteter, samt at forskningsinstitutionerne fik bedre muligheder for at inddatere data i systemet.

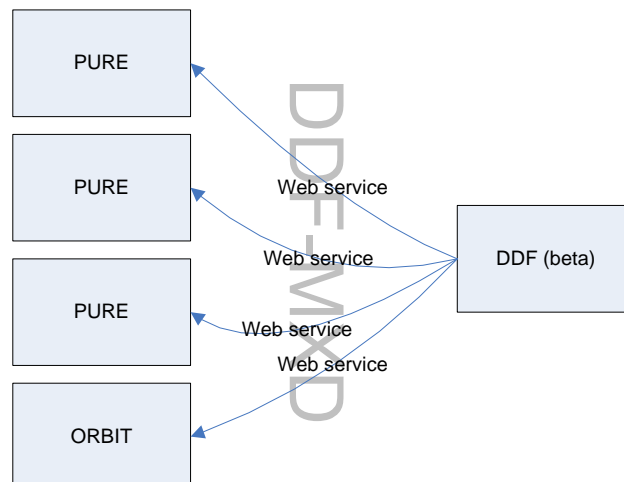
Basen samler oplysninger om forskningspublikationer, forskningsprojekter og forskningsinstitutioner. Hidtil har det især været publikationsopsamlingen, der har været mest dækkende, og frem til opsamlingen af data blev sat i stå i forbindelse med forberedelsen af videreudvikling var der opsamlet ca. 215.000 titler i databasen.

DDF får data fra universiteterne, samt en lang række organisationer, bl.a. RISØ og Kræftens Bekæmpelse.

Der er i samarbejde mellem universiteterne og DEFF udviklet et fælles XML-skema (DDF-MXD). Dette skema anvendes som udvekslingsformat mellem PURE, Orbit og DDF.

DDF er under videreudvikling og findes i dag i en beta-udgave, som endnu ikke er sat i drift. Denne beta-udgave, som af DTV betegnes som klar til drift rent teknisk. Denne beta-udgave indeholder en datahøster, som via web service interfaces i både PURE og Orbit installationerne høster data til DDF.

Integrationen er illustreret i nedenstående model:



Leverandørforhold

DDF er i dag en del af Danmarks Elektroniske Forskningsbibliotek, som drives af Kulturministeriet, Undervisningsministeriet og Ministeriet for Videnskab, Teknologi og Udvikling.

Driftsmodel

Den løbende drift, vedligeholdelse og videreudvikling af DDF forestås af DTV på en aftale med DEFF.

Workflow og datakvalitet

Data i DDF høstes – I den beta-udgave som forventes sat i drift snarest - fra universiteternes PURE- og Orbit-installationer via et web services interface.

Teknisk arkitektur

DDF er udelukkende bygget på open source teknologier og den kode som er udviklet til DDF er ligeledes open source. DDF er baseret på en MySQL database og funktionaliteten er kodet i Java. Det gælder også den datahøster som høster data fra PURE-systemerne og Orbit via et web services interface.

3.4 Samlet vurdering af det eksisterende systemlandskab

Gartner konstaterer, at PURE og Orbit begge er systemer, som er implementeret og i brug til forskningsregistrering på de danske universiteter. Længden af erfaringerne er forskellige fra universitet til universitet, men begge systemer har demonstreret at de kan fungere til at understøtte registreringen af forskningspublikationer.

Både PURE og Orbit har ligeledes en sund arkitektur bygget på tidssvarende teknologier.

Med dataudvekslingsstandarden DDF-MXD er høstning af data en forholdsvis simpel opgave. Den tekniske platform til DDF vurderes at kunne løfte opgaven med indhøstning af registreringer på betryggende vis.

Man må derfor samlet set konstatere, at der eksisterer en række systemer og en integrationsmodel, som med en række mindre tilføjelser kan tjene som grundlag for en bibliometrisk forskningsindikator.

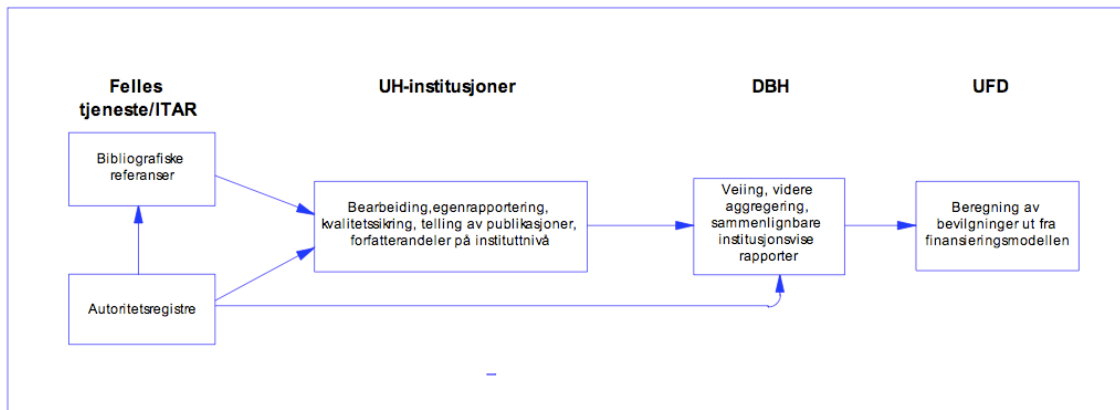
4.0 Internationale erfaringer

4.1 Norge

Norge har gennem en årrække arbejdet på at indføre kvantitative mål for forskningspublikation som basis for tildeling af forskningsmidler, og en del af midlerne tildeles i dag ud fra publikationsstatistik. Systemet for tildeling af midler blev beskrevet i 2004 og taget i brug første gang i 2005.

I Norge anvendes der primært to systemer til forskningsregistrering: ForskDok og Frida. Der er etableret en fællestjeneste, som stiller autoritetsregistre til rådighed, og forestår høstningen af data fra registreringssystemerne. Den tekniske del af fællestjenesten hedder ITAR.

Det er de enkelte institutioners ansvar at levere korrekte data til DBH (DataBase for Høgre utdanning), som forestår udarbejdelsen af statistikker over de enkelte institutioners andel af den samlede publikationsmasse.



ITAR

Sammen med forskningsregistreringssystemerne, er systemet ITAR (Importtjeneste og Autoritetsregistre) udviklet til dannelse af autoritative lister over videnskabelige tidsskrifter og forlag. ITAR benytter ISI, Bibsys og Norart som grundlag for dannelsen af listerne. Bibsys og Norart er registre over norske, videnskabelige tidsskrifter, som ikke indgår i ISI. De autoritative lister opdateres månedligt og leveres til de institutioner som anvender ForskDok og Frida, hvor data dernæst gøres tilgængelige i forbindelse med indtastning.

ITAR indeholder i dag omkring 14.000 titler på videnskabelige tidsskrifter, bøger o.lign., samt ca. 450 forlag. Det er væsentligt at bemærke, at ITAR ikke kun er lister over A-tidsskrifter, men tilsigter at rumme alle potentielle publikationskanaler for norske forskere, nationale som internationale. Inden for rammerne af ITAR eksisterer der et rangordnings-system, hvor tidsskrifter inddeles i 3 kategorier: "2" (A-tidsskrifter), "1", eller "andre". Publikationskanaler kan foreslås af alle brugere af systemet, og det afgøres af den centrale fællestjeneste, om publikationen kan optages i registret, og om den skal tildeles kategorien "1" eller "andre". Kategori "2" tildeles i samråd med nedsatte, nationale fagråd. Kun tidsskrifter ranked 1 og 2 medtælles i tildelingen af midler.

Eftersom midler tildeles på baggrund af institutionernes andel af den samlede publikationsmasse (vægtet efter publikationskanalernes rangorden) er det væsentligt at sikre korrekte oplysninger om de enkelte forfatters institutionelle tilhørsforhold. I mangel af et centralt register, håndteres det i dag på den måde, at det er op til institutionerne selv at sikre at de opgivne tilhørsforhold er korrekte. Oplysninger om forfattertilhørsforholdet er således

ikke en del af fællestjenestens ydelser. Man har i Norge ikke vurderet at det var praktisk gennemførligt at udarbejde et sådant register.

Frida

Frida (Forskningsresultater, Information og Dokumentasjon av vitenskapelige Aktiviteter) udvikles i et samarbejde mellem de fire norske universiteter, med henblik på at skabe så gode rammer som muligt for universiteternes forskningsregistrering.

Systemet indeholder et rapporteringsmodul, der tillader udtræk af bl.a. publikationsstatistik for de enkelte forskningsinstitutioner.

Frida indeholder også et forskerkatalog, hvor de enkelte forskere selv kan oprette profiler. Forskerkataloget er tænkt som et redskab for såvel den videnskabelige verden, som for offentligheden, for hvem den skal bidrage til at skabe med at skabe overblik over videnskabelig ekspertise og kompetencer. Forskerkataloget er dog ikke komplet, og den anvendes heller ikke i statistikker over forskningsaktiviteten.

Det er hensigten af Frida med tiden skal udvikles til også at fungere som en portal for forskningsformidling vendt mod den almindelige offentlighed og erhvervslivet, på samme måde som DDF i Danmark. Til det formål skal der udvikles et projektmodul, som er under planlægning.

ForskDok

ForskDok udvikles af BibSys (der svarer til det danske DanBib), og leverer foruden ForskDok også bibliotekssystemer, samt dokument- og billedarkiveringsystemer.

ForskDok anvendes af over 50 universiteter/høgskoler, og er under fortsat udvikling. Det er planlagt at samtlige universiteter skal overgå til at benytte Frida-systemet, men dette er endnu ikke gennemført.

ForskDok har både et modul til registrering af publikationer (ForskPub), og et projektarkiv (ForskPro). Ansvaret for at de inddaterede oplysninger er korrekt forvaltes lokalt. Dette sker efter lokale retningslinier.

ForskDok muliggør central indsamling af data, som stilles til rådighed for søgning i Bibsys Ask-systemet, ligesom registrerede brugere har adgang til at udtage statistikker om inddaterede publikationer fra de tilknyttede institutioner.

4.2 England

En central registrering af forskningspublikationer er under planlægning i England. Den væsentligste bevæggrund er i modsætning til f.eks. Norge ikke at indføre en bevillingsstruktur baseret på kvalitetsvurdering af forskningsindsatsen, men at ændre den eksisterende, Research Assessment Exercise (RAE). Det eksisterende system er baseret på et 6-årligt tilbagevendende review, hvor ekspertpaneler har vurderet forskningskvaliteten på alle institutioner indenfor diverse forskningsområder.

Denne proces har vist sig at være umådelig arbejdskrævende for institutionerne, der skal dokumentere deres arbejde i rapporter o.lign., og er tillige blevet kritiseret voldsomt for at have uigennemskuelige kvalitetskriterier, hvilket har ført til et politisk ønske om at skabe en proces, der er lettere at administrere, og som baserer sig på transparente principper for kvalitetsvurderingen.

I forbindelse med RAE 2008 planlægges det at gennemføre et pilotforsøg, hvor en publikationsbaseret scoring af forskningsinstitutionerne udarbejdes parallelt med den planlagte reviewproces.

Dataindsamlingen til RAE 2008 er endnu ikke afsluttet, og der er ikke offentliggjort detaljerede planer for pilotforsøgets gennemførelse endnu.

4.3 Holland

I Holland har Metis været anvendt siden 1993 som forskningsregistreringssystem. Det er implementeret på alle 13 hollandske universiteter.

Systemet anvendes dog ikke som en del af bevillingsprocessen til de hollandske universiteter og systemerne fungerer som udgangspunkt som isolerede installationer på de enkelte universiteter.

4.4 Opsamling af internationale erfaringer

Anvendelsen af statistik over publikationer anvendes i vidt omfang på universiteterne internationalt til intern styring, til bevilling af projektmidler og ved ansættelser af forskere.

Det er dog ifølge Gartners erfaringer kun i Norge, at man har erfaringer med nationalt at koble en national bibliometrisk forskningsindikator til bevillingsprocessen til Universiteterne.

Det er derfor naturligt at lære af erfaringerne i Norge med hensyn til at skabe en national forskningsindikator.

I Norge har man ligesom i Holland ét system til registrering af forskning på tværs af universiteterne. I Holland driftes systemerne decentralt som i Danmark. I Norge driftes de centralt, men de enkelte universiteters databaser er isoleret fra hinanden.

5.0 Rammer for en dansk forskningsdatabase

Gartners analyse af opgaven med at registrere forskningsaktivitet har ført til følgende overordnede observationer, som bør danne rammer for den fremtidige løsning, herunder designkriterier for systemet.

- Registreringen af forskningen er et vigtigt værktøj internt på universiteterne og anvendes / skal anvendes både til styring og til formidling. Derfor skal det system som anvendes til registrering af forskning både understøtte den bibliometriske forskningsindikator og anvendelser internt til styring og formidling på universiteterne.
- Datakvaliteten er det mest kritiske element ved systemet. Da data primært er baseret på menneskelig indtastning, vil der kontinuerligt optræde fejl i data.
- Forskernes indberetningsgrænseflade skal have en høj tilgængelighed, men kravene til tilgængelighed er ikke som i forretningskritiske systemer.
- Undersøgelser på DTV har vist, at det ved validering er realistisk at forvente fejl i 1 – 3 % af records. Det er således ikke muligt at forestille sig 100% korrekte data i en database for forskningsregistrering.
- Erfaringerne med at registrere forskningspublikationer er meget forskellig på universiteterne. Der er universiteter som har arbejdet med det en årrække og universiteter som lige er gået i gang og endnu ikke har det først fulde års forskningsregistreringer.
- Erfaringerne fra de som har indført systemerne siger at det er en omfattende proces at implementere hos forskerne. Derfor bør det system, som skal understøtte den bibliometriske forskningsindikator i første omgang præsentere en næsten uforandret brugergrænseflade overfor slutbrugerne.
- Der er ikke gode samlede tal for fejlniveauet i de data som høstes ind i den nye beta-version af DDF, men den politiske bevågenhed gør det fornuftigt at undersøge mulighederne for at højne datakvaliteten yderligere.

Styringsmæssige rammer

- Det er nødvendigt at universiteterne har en form for ejerskab over systemerne til registrering af forskning. Det er således en forudsætning at forskerne kun registrerer publikationer ét sted, som det sker i dag.
- Det er nødvendigt at have central styring af de autoritative lister og de bør stilles til rådighed for universiteterne.
- Det er nødvendigt at etablere et vist niveau af validering centralt. Det vil i sær være nødvendigt i en indkøringsfase.

5.1 Rammer udstukket af det faglige udvalg

Det faglige udvalg, som er nedsat til at udvikle en model for vurdering af forskningsaktivitet har allerede defineret overordnede principper for den bibliometriske forskningsindikator, som danner ramme for den måde IT-løsningen skal fungere på.

Principperne for den bibliometriske forskningsindikator betyder:

- At der vil blive behov for deduplicering af tidsskrifter. Indikatoren bliver sådan sammensat, at artikelsamarbejder mellem forskere fra flere danske universiteter skal deles mellem universiteterne. Doubletter som er fremkommet ved at forskere fra begge institutioner indberetter den samme artikel i den nationale forskningsdatabase skal derfor fjernes.

- Der vil blive indført autoritative publikationslister. Det vil sige at indikatoren vil blive baseret på en liste af tidsskrifter, forlag og evt. også konferencer. Hvis en publikation er foretaget gennem en af disse kanaler, vil den tælle med i indikatoren. Denne autoritative liste bør derfor implementeres i en database, som er tilgængelig, når forskeren registrerer publikationer.

5.2 Tidsmæssige bindinger for etableringen af en fælles forskningsdatabase

Det er målsætningen at kunne anvende den bibliometriske forskningsindikator i forbindelse med tildeling af bevillinger til universiteterne for 2010 på basis af tal fra publiceringer i 2008. Denne målsætning medfører nogle meget markante bindinger for etableringen af den nationale forskningsdatabase.

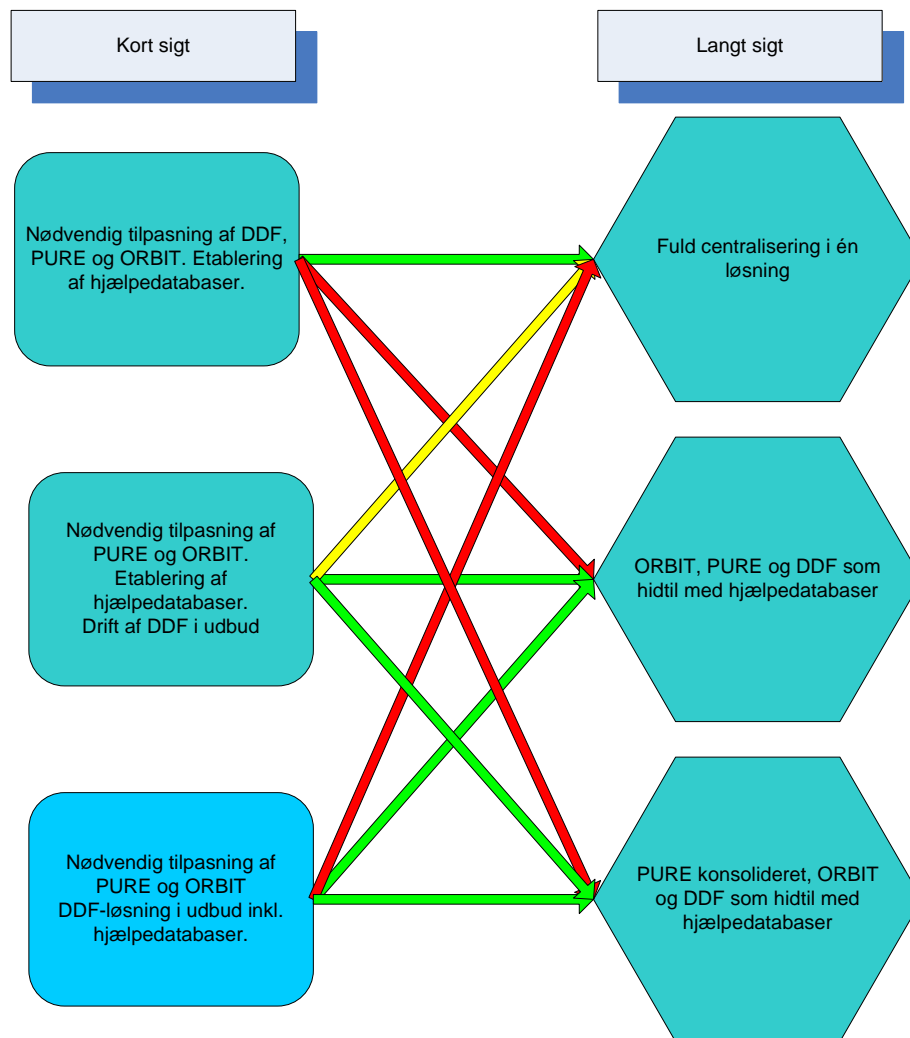
Tiden udgør en vigtig faktor i forbindelse med definitionen af de scenarier, som kan opstilles for driften af den nationale forskningsdatabase.

- Det er forventningen at den autoritative liste over publikationskanaler (tidsskrifter, bogforlag, etc.) ligger klar marts 2008.
- For at give tid nok til at brugerne kan indtaste publikationer og valideringer kan finde sted lokalt på universiteterne, er det nødvendigt at have indtastningssystemer til rådighed tidligt i 2008
- Hjælpedatabaserne må være tilgængelige til registrering i PURE og Orbit marts 2008.
- Forskningsregistreringen for 2008, som senest skal foregå i efteråret 2008 må foregå i de eksisterende PURE og Orbit systemer.
- Hjælpedatabaserne fungerer ikke før der er lavet mindre tilpasninger til PURE og Orbit.
- Det er derfor ikke realistisk at samle registreringer til scoring for 2008 på baggrund af et udbud.

6.0 Fremtidige modeller for en national forskningsdatabase

De rammer og forudsætninger som er ekspliciteret vedrørende etableringen af en bibliometrisk forskningsindikator giver en række handlingsmuligheder på helt kort sigt og på længere sigt i forhold til etableringen af et system som kan producere den bibliometriske forskningsindikator og på samme tid understøtte universiteternes interne anvendelse af disse data til styring og formidling.

I modellen nedenfor er illustreret handlingsmulighederne på kort sigt og langt sigt samt de indbyrdes sammenhænge.



Langt sigt

På langt sigt kan man bevæge sig mod to mål:

- at etablere en fuldt centraliseret løsning, som dækker den funktionalitet som har ligget i PURE, Orbit og DDF med ét system,
- eller, at bibeholde de eksisterende systemer.

Den centrale model kræver den største investering men giver en række IT-driftsmæssige forenklinger og også mindre procesmæssige forenklinger.

Hvis man vælger den decentrale model, kan man evt. realisere nogle af de IT-driftsmæssige forenklinger ved at konsolidere driften af de mange PURE installationer.

Kort sigt

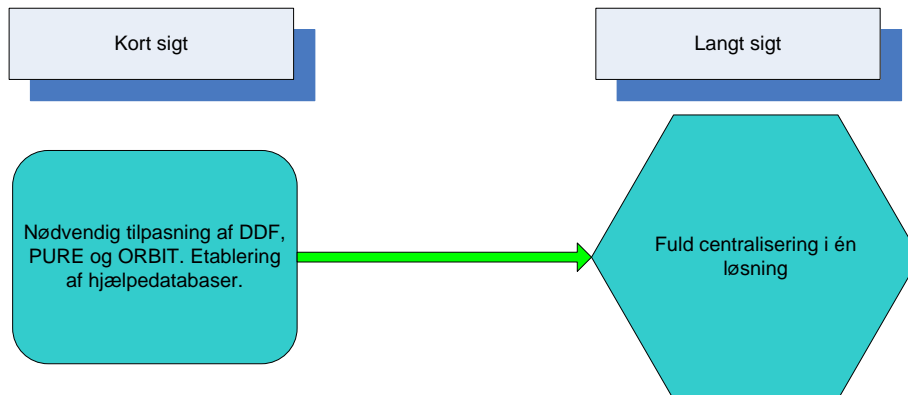
På kort sigt kan man ikke etablere en central løsning og her står man overfor at vælge mellem :

- Fortsat at lade DTV drifte det den centrale del af forskningsdatabasen, baseret på teknologien på den nuværende betaversion af DDF.
- Udbyde driften af systemet baseret på den nuværende betaversion af DDF.
- Sende etablering og drift af en løsning tilsvarende betaversionen af DDF i udbud.

Gartner kan imidlertid ikke finde nogen gode argumenter for, hvorfor man skal sende etablering og drift af et tilsvarende system i udbud da det som findes kan udgøre en glimrende platform, hvor udviklingsomkostningerne er afholdt og afskrevet.

Derfor står man overfor et valg mellem to scenarier, som præsenteres og vurderes nedenfor.

6.1 Scenarie 1: fuld centralisering



Beskrivelse

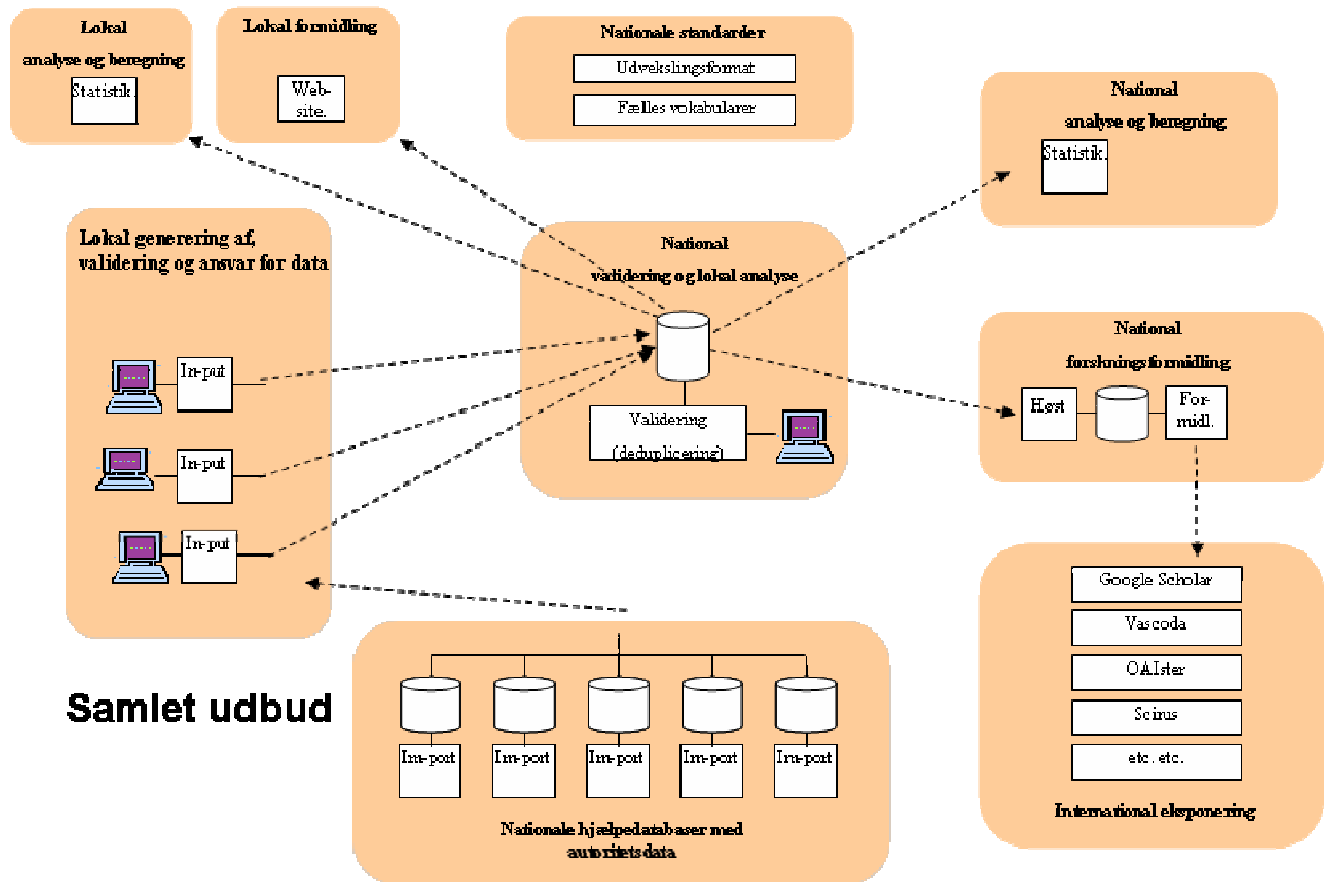
Her gennemfører man på kort sigt en mindre række tilpasninger på kort sigt, som gør det muligt at højne datakvaliteten i systemet i den periode, hvor man endnu ikke har etableret en central løsning.

Parallelt med disse aktiviteter identificerer man en organisation som kan varetage specifikation, indkøb og drift af én central løsning, som kan varetage det som er specificeret i afsnittet "Krav til en centraliseret løsning".

Det er nødvendigt at universiteterne har ejerskab af en sådan centraliseret løsning og UNI-IT kunne danne en naturlig organisatorisk ramme for systemet, med styregruppedeltagelse fra Forsknings- og Innovationsstyrelsen.

Målarkitektur

Målarkitekturen for det samlede systemkompleks kommer i en centraliseret løsning til at se ud som den nendestående tegning:



Vurdering

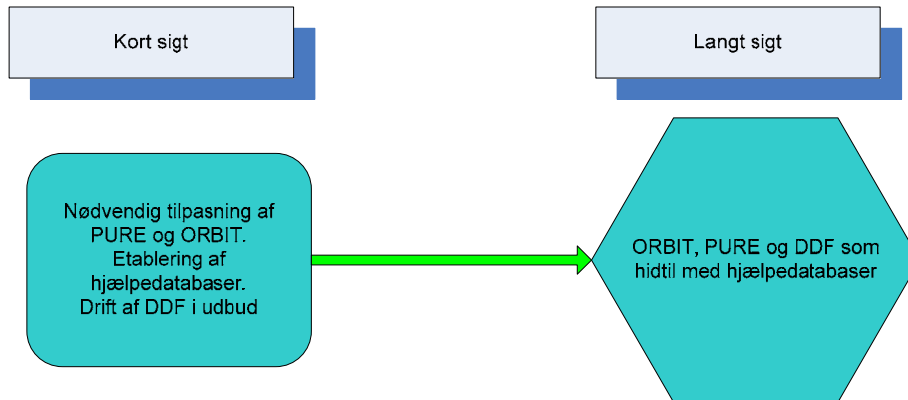
Fordele:

- Laveste omkostninger på langt sigt til drift og vedligeholdelse.
- Fjerner behovet for deduplicering, da dubletter vil blive fanget ved indtastningen.
- Giver mulighed for at effektivisere investeringer i valideringsteknologi og processer ét sted frem for decentralt på universiteternes biblioteker.

Ulemper

- Centraliseringen kræver en større investering end scenarie 2 og kan ikke gennemføres på kort sigt.
- Vil kræve at brugere lærer ny grænseflade.
- Ufleksibel styringsmodel som alt andet lige vil give universiteterne en mindre fleksibel mulighed for lokal tilpasning. Der kan dog kompenseres i vidt omfang gennem etablering af fleksible grænseflader i systemet.

6.2 Scenarie 2: Decentral model



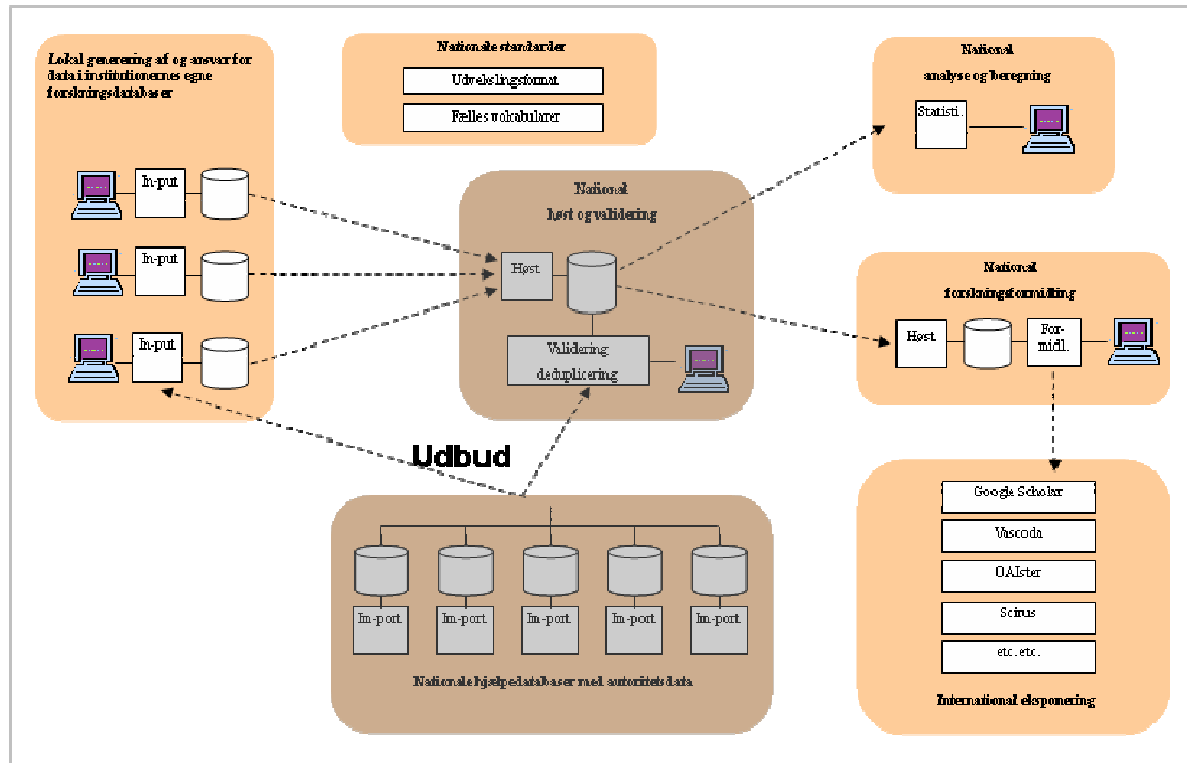
Beskrivelse

Her foretager man på kort sigt de nødvendige tilpasninger til PURE og Orbit og forbereder samtidigt et udbud på driften af den nationale forskningsdatabase, baseret på det system som i dag udgør betaversionen af DDF.

Scenariet indebærer at man på sigt etablerer en integration mellem den nationale forskningsdatabase og PURE og Orbit, så rettelser til registreringer som er identificeret centralt kan routes tilbage til brugeren af PURE og Orbit, så de kan indføje eller godkende rettelsen.

Det er i princippet muligt at forestille sig scenarie 2 gennemført uden at udbyde driften af den nationale forskningsdatabase og køre videre på den aftale som eksisterer mellem DEFF og DTV. Det er dog Gartners vurdering, at DTV uvægerligt vil ende i interessekonflikter ved både at være en del af DTU og samtidigt skulle repræsentere interesser knyttet til det bibliometriske forskningsindeks.

Målarkitektur



Vurdering

Fordele:

- Scenariet kan gennemføres med få nødvendige ændringer og med relativt små investeringer.
- Sikrer at brugerne ikke mærker nogle systemmæssige ændringer og med minimale justeringer kan bevare den brugergrænseflade som kendes i dag..
- Sikrer universiteterne fleksibilitet i forhold til interne anvendelser af forskningsregistreringer.

Ulemper

- Høje samlede omkostninger på langt sigt i form af et urealiseret besparelspotentiale.
- Kræver udbygning af integrationen mellem den centrale database, PURE og Orbit for at kunne håndtere valideringsworkflow.
- Betyder at dedupliceringsprocessen skal varetages centralt. Denne proces forsvinder i den fuldt centraliserede model.

7.0 Validering og etablering af hjælpedatabaser

At opnå en datakvalitet som er acceptabel, er det mest kritiske succeskriterium for det system som skal skabe den bibliometriske forskningsindikator.

Uanset hvilket af de to scenarier som vælges, er det anbefalingen, bl.a. på baggrund af erfaringer hos DTV, at man etablerer en række hjælpedatabaser for at bidrage til at højne datakvaliteten. En hjælpedatabase kan f.eks. være et subset af ISI eller den autoritative liste over publikationskanaler.

Hjælpedatabaserne kan indsættes i forbindelse med valideringen¹ på to måder:

1. Ved at registreringssystemet laver en on-line validering af forskerens indtastning af publikationens titel eller tidsskriftets titel.

Denne metode er at foretrække, fordi den giver mulighed for at forskeren kan godkende hjælpedatabasens forslag til indsætning af data. Metoden virker dog kun, hvis publikationen findes registreret i hjælpedatabasen allerede til det tidspunkt hvor forskeren vælger at registrere publikationen. Derfor vil det bidrage til datakvaliteten også at gennemføre en mere kompleks validering,

2. ved at man periodisk laver batch-kørsler som foretager en matchning mellem en hjælpedatabase og databasen over registreringer.

Denne matchning vil dels kunne identificere publikationer, som ikke er registrerede, dels identificere oplysninger om en publikation, som ikke stemmer overens. Ved identifikation af en sådan diskrepans kan forskeren notificeres og man kan bede vedkommende om at registrere publikationen eller rette i den eksisterende registrering.

I nedenstående skema findes en oversigt over potentielle hjælpedatabaser og Gartners vurdering af, hvor vidt det er fornuftigt at investere i dem. Vurderingen er baseret på en vægtning af investeringens størrelse i forhold til dens forventede effekt på datakvaliteten.

Hjælpedatabase	Beskrivelse	Vurdering	
Database over publikationer.	Der etableres en database med et subset af f.eks. ISI eller Scopus.	Giver både mulighed for on-line og batch validering. Erfaringer siger at metoden fanger knap halvdelen af publikationerne	
Oversigt over tidsskrifter	Oversigt over tidsskrifter med bl.a. angivelse af status (peer-reviewed etc.) samt øvrige bibliometriske data som f.eks. udgivelsesland	Databasen befolkes centralt på baggrund af fagudvalgenes arbejde med proces til at forskere kan foreslå tidsskrifter. Er let at etablere og forhindrer bl.a. tastefejl.	
Oversigt over konferencer	Oversigt over konferencer med bl.a. angivelse af status	Databasen befolkes centralt på baggrund af fagudvalgenes arbejde med proces til at forskere kan foreslå	

¹ Validering benyttes her både i den betydning, hvor systemet mens indtastningen foregår validerer en indtastning ved f.eks. at slå op i en referenceliste og senere check af, hvor vidt de indtastede data er korrekte.

Hjælpedatabase	Beskrivelse	Vurdering	
	(peer-reviewed etc.) samt øvrige bibliometriske data som f.eks. udgivelsesland	nye konferencer. Det er ikke endeligt afklaret om publikationer til proceedings fra konferencen bliver en del af den autoritative liste og hjælpedatabasen er kun relevant, hvis det er tilfældet. Er let at etablere og forhindrer bl.a. tastefejl.	
Oversigt over bogforlag	Oversigt over bogforlag samt øvrige bibliometriske data som f.eks. udgivelsesland	Databasen befolkes centralt på baggrund af fagudvalgenes arbejde med proces til at forskere kan foreslå nye bogforlag. Er let at etablere og forhindrer bl.a. tastefejl.	
Central persondatabase	En løsning så der centralt findes en fortegnelse over hvor forskere er ansat i hvilke perioder.	Hjælpedatabasen giver mulighed for at validere at en forfatters organisatoriske tilhørsforhold er korrekt. Denne løsning vil være meget kompleks at etablere og vedligeholde. Den vil bl.a. kræve integration med universiteternes personalesystemer.	

Farvekoder:

Grøn = det anbefales etableret af Gartner

Rød = anbefales ikke etableret af Gartner

Hvilke(n) database(r) over publikationer

Gartner anbefaler at man etablerer en hjælpedatabase, som indeholder en fortegnelse over publikationer med bidragsydere, som er ansat ved danske universiteter. Fortegnelsen skal hentes fra en eller flere af de internationale fortegnelser over publicerede videnskabelige artikler.

Det er Gartners vurdering, at man bør tage udgangspunkt i enten ISI eller Scopus ved etableringen af hjælpedatabasen.

På et senere tidspunkt bør det undersøges om man kan foretage validering i form af online søgninger i Google Scholar eller Windows Live academic i forbindelse med forskerens indtastning af data, men det vurderes på nuværende tidspunkt som en løsning, der er kompleks at etablere og som ikke giver væsentligt flere resultater end ISI eller Scopus.

Database	Vurdering
ISI	Vurderes umiddelbart anvendelig. DTV anvender allerede batch-validering op i mod et subset af ISI.
Scopus	Den overordnede vurdering er at Scopus er lige så godt dækkende som ISI på naturvidenskab, og bedre på samfundsvidenskab og humaniora. Scopus er lidt svagere på historiske data, hvilket er irrelevant i denne sammenhæng.

Database	Vurdering
Ejos	Biblioteksløsning der leveres af Elsevier. Kan levere publikationsdata, som opbevares lokalt. Specifikt designet til integration med andre systemer. Skal suppleres af ISI/Scopus for at løse den forestående opgave.
Google Scholar	Indeholder i modsætning til ISI/Scopus også bøger. Giver ikke umiddelbart mulighed for at etablere et lokalt subset af data og on-line søgninger vurderes som komplekst at gennemføre på nuværende tidspunkt.
Windows Live Academic	Indeholder i modsætning til ISI/Scopus også bøger og konferenceproceedings. Giver ikke umiddelbart mulighed for at etablere et lokalt subset af data og on-line søgninger vurderes som komplekst at gennemføre på nuværende tidspunkt.
CiteSeer	Begrænser sig primært til informationsvidenskab og er dækket af ISI/Scopus

7.1 Decentral og central validering

Det er sandsynligt at man i forbindelse med indførelsen af den bibliometriske forskningsindikator vil se at universiteterne af egen drift vil intensivere valideringen. Der er væsentlig forskel på de valideringsprocesser som sker på de forskellige universiteter og det vil være hensigtsmæssigt at man enes om nogle fælles processer eller i det mindste principper for valideringen.

Validering sker med to formål:

- At sikre sig at flest mulige publikationer bliver registreret.
- At sikre sig mod fejl.

Den første type validering vil universiteterne uden tvivl foretage og de vil her i et vist omfang blive hjulpet af den hjælpedatabase som etableres medbset af ISI eller Scopus. Disse valideringer bør ske hos universiteterne uanset om man vælger scenarie 1 eller 2.

Centralt vil man primært foretage validering af korrekthed af indtastningerne og især er dedupliceringsopgaven vigtig og kan kun løses centralt.

Valideringsprocesser

I det følgende listes en række valideringsprocesser, som er inspireret af de processer som er defineret og beskrevet for Frida i Norge. Det skal i den forbindelse bemærkes, at Gartner anbefaler at den endelige beslutning omkring hvilke valideringsprocesser, som besluttet bør bero på en biblioteksfaglig vurdering. Desuden vil processerne skulle justeres i takt med at man får erfaringer med, hvor de faktiske problemer med datakvalitet opstår.

De processer, som er foreslået er i Gartners optik de processer, som vil have den største virkning på datakvaliteten. Dog skal det i den forbindelse bemærkes, at den mest basale og effektive måde at højne datakvaliteten på, er at sørge for at forskere har en interesse i at registrere forskningen korrekt, f.eks. fordi den influerer på den evaluering forskeren får hos universitetsledelsen.

Der eksisterer ikke fælles processer i dag til validering på tværs af forskningsinstitutionerne og det er Gartners vurdering at der ikke er et absolut behov for at processerne er harmoniseret, så længe universiteterne har procesvarianter som giver de samme effekter på datakvaliteten.

Navn	Beskrivelse	Institution	Centralt
Enheder	Superbruger/administrator sikrer at alle forskere i enheden er registreret i systemet. Det sker i dag typisk via en integration til institutionens personalesystem.	X	
Indtastning	Brugergrænsefladerne sørger for validering af de data som valideres i dag i PURE og ORBIT plus opslag i de ovenfor nævnte hjælpedatabaser	X	
Forfattercheck	Forfatteren laver i forbindelse med indberetningen et manuelt check af om oplysningerne er korrekte.	X	
Enhedscheck	Superbruger/administrator laver manuelt check for forskningsenheden og sørger for evt. udsendelse af remindere til forskere, som ikke har registreret.	X	
Batchkørsel – institution	Batchkørsel i hjælpedatabase for at identificere publikationer som ikke er registreret i den nationale forskningsdatabase. Dette understøttes i dag i Orbit og kan i første omgang udføres manuelt i PURE. ¹	X	X
Institutionscheck	Superbruger/administrator laver manuelt check af dubletter for institutionen.	X	
Central deduplicering	Check af dubletter i den nationale forskningsdatabase og notifikation af de involverede forskere, hvis der ikke bare er tale om en simpel dublet.		X
Stikprøver	Udtagelse af stikprøver for at vurdere datakvaliteten.		X

IT-understøttelsen af disse processer er efter Gartners vurdering tilstrækkelig i både PURE og Orbit til at det er forsvarligt at etablere den bibliometriske forskningsindikator. Ved en fuld centraliseret model vil der dog være et potentiale for at forbedre understøttelsen af de ovenstående processer. Det samme gælder i det decentrale scenarie.

Det gælder bl.a. at PURE's integration med hjælpedatabaserne kan udbygges, så det både understøtter at forskningsregistreringer oprettes via import fra hjælpedatabasen (baseret på ISI eller SCOPUS) og via validering når forskeren registrerer publikationen.

¹ Hvis man vælger det decentrale scenarie, bør denne investering foretages i PURE. Hvis man vælger det centraliserede scenarie, bør investeringen ikke foretages i PURE.

Fraktionering af artikler og bøger

I den bibliometriske forskningsindikator vil der sandsynligvis (det er ikke besluttet ved redaktionens slutning) blive udført fraktionering af artikler. Det vil sige, at artikler som har flere forfattere fra flere universiteter, skal fraktioneres, så der gives f.eks. halve point til et universitet, hvis der er tale om en artikel med to forfattere fra to forskellige universiteter.

Hvis denne fraktionering vedtages, vil det være centralt at udføre grundig validering af de artikler, som har flere forfattere.

I første omgang anbefales det, at der sker en manuel central validering af artikler med flere forfattere og at der sker et check af fraktioneringen i forbindelse med dubletter, som skyldes at to forskere fra to universiteter har registreret den samme artikel, som de er fælles om.

Hvis der skal udføres fraktionering i forhold til artikler og bøger som er forfattet i samarbejde med udenlandske forfattere, vil det komplicere fraktioneringen væsentligt. I praksis vil man her kun have mulighed for at validere de artikler som kan identificeres via ISI/Scopus.

Støtteprocesser

Det er besluttet at indføre autoritative lister over hvilke tidsskrifter og hvilke forlag som tæller i den bibliometriske forskningsindikator. Der er på baggrund af den behov for en proces for forskere eller institutioner foreslår nye tidsskrifter og forlag til listen. Disse forslag kan med tiden understøttes systemmæssigt men kan i første omgang baseres på almindelig e-mail korrespondance.

8.0 Aktiviteter på kort sigt

Uanset hvilket scenarie som man vælger at arbejde sig hen i mod, er der nogen tilpasninger til de eksisterende systemer, som på kort sigt skal løses. Disse opgaver skal løses for at kunne leve op til målet om at kunne udarbejde en måling baseret på publiceringer foretaget i 2008.

Tilpasningerne er nødvendige for at sikre en god datakvalitet. Ud fra princippet om at valideringer af data sker bedst ved indtastningen, anbefaler Gartner følgende tilpasninger, som kan sættes i gang med det samme:

- Udarbejd en løsning, som centralt stiller den autoritative tidsskriftsliste og liste over godkendte forlag til rådighed for opslag for PURE og Orbit. Det er Gartners anbefaling at det stilles til rådighed som en web service.
- Udarbejd en løsning som stiller en relevant delmængde af ISI eller Scopus til rådighed for opslag fra PURE og Orbit. Det er Gartners anbefaling at det stilles til rådighed som en web service.
- Tilpas PURE og Orbit, så de felter som udfyldes vedrørende titel på artikel (via ISI/Scopus), samt tidsskriftet/bogforlaget/konferencen (via den autoritative liste) valideres under indtastningen ved at anvende de web service som stilles til rådighed.
- Indfør de valideringsprocesser som er nævnt i afsnit 7.2, i det omfang de ikke allerede er på plads på universiteterne.
- Hvis den bibliometriske forskningsindikator kræver en tilpasning af udvekslingsformatet, kræver det tilpasning af systemerne. Ved redaktionens slutning er disse krav ikke på plads. Det er dog vurderingen at ændringerne vil blive relativt små.

9.0 Opgaver og organisering af en central organisation

Det er Gartner klare vurdering, at der uanset scenarie er behov for at etablere en central styregruppe med deltagelse fra både universiteter og F&I. Denne styregruppe bør på samme tid være faglig og teknisk styregruppe, dvs. behandle både spørgsmål vedrørende beregningen af publiceringsaktiviteten samt spørgsmål vedrørende det system, som indsamler data og udarbejder statistikken.

Desuden er der behov for en central organisation, som kan varetage følgende opgaver:

- Drift af IT-system/styring af driftsleverandør.
- Vedligehold af den nationale forskningsdatabase.
- Tilpasninger og videreudvikling af systemet / styring af underleverandører.
- Vedligeholdelse af den autoritative liste og andre hjælpedatabaser.
- Udførelse af valideringsprocesser som beskrevet i afsnit 7.1.

Det er Gartners vurdering at en sådan driftsorganisation vil kunne varetages af ca. 2 årsværk i scenarie 2, forudsat at driften og større IT-udviklingsopgaver leveres af en underleverandør. Det er også forudsat at man kan finde den rette kompetencesammensætning eller at man fordeler det over flere personer i en organisation, hvor man kan opnå synergier.

I et centralt scenarie (1) vil der være behov for flere ressourcer til også at vedligeholde den del af systemet som er vendt mod brugerne, samt yderligere ressourcer til at varetage den validering som foregår i dag på universiteterne.

Til sammenligning har man i den organisation, som står bag FRIDA i Norge 2,5 årsværk til hvad der benævnes sekretariatsfunktion og brugerstøtte. Hertil kommer 3 årsværk til udvikling af FRIDA.

10.0 Økonomiske betragtninger

I det følgende er foretaget retningsgivende grovestimater for forskellige dele af de initiativer som er nævnt i rapporten:

10.1 IT-initiativer på kort sigt

Uanset om man vælger fuld centralisering eller et decentralt scenarie vil man for at kunne levere en bibliometrisk forskningsindikator for publikationer i 2008 være nødsaget til at lave en række kortsigtede tilpasninger til PURE, Orbit og den nationale forskningsdatabase, for at den kan fungere:

- Etablering af hjælpedatabaser: autoritative lister og subset af ISI eller Scopus. Dette bør kunne implementeres i den nationale forskningsdatabase for under 500.000,- i det omfang den ikke kan indgå som en del af den eksisterende forvaltningsaftale mellem DEFF og DTV.
- Tilpasning af PURE og Orbit til så de kan udnytte hjælpedatabaserne. Det vurderes at det ikke bør koste mere end 50.000,- pr. institution i det omfang det ikke er indeholdt i de eksisterende vedligeholdelsesaftaler med leverandøren.
- Evt. tilpasning af udvekslingsformatet. Dette kan ikke estimeres på nuværende tidspunkt, da kravene til den bibliometriske forskningsindikator ikke ligger fast. Det vurderes dog at udgøre en begrænset omkostning.

10.2 Investeringer decentralt vs. centralt scenarie

I det følgende er estimeret de investeringer, som vurderes nødvendige for at realisere det decentrale og det centrale scenarie på sigt:

I det fuldt centraliserede scenarie vil man skulle gennemføre et udbud på et sådant system. For et sådant udbud vil der være mindst én leverandør med et system, som med modifikationer vil kunne anvendes som svar på udbuddet (Atira A/S). Desuden gælder det, at den kode som kører Orbit og DDF er open source kode, hvilket bør betyde at en leverandør som formår at udnytte dette effektivt, vil kunne afgive et konkurrencedygtigt tilbud.

Ud fra et indledende grovestimat, vurderes det muligt at få en systemintegrator til at udvikle en central løsning som den der er skitseret i 11.0 for ca. 2 mio. kr.¹ Dette estimat er baseret på, at det meste af funktionaliteten i et centralt system ligger fast (her er tale om et system, hvis specifikationer i høj grad ligger fast).

Det centraliserede scenarie afføder desuden en række omkostninger hos de universiteter som anvender forskningsregistreringerne til formidling. Der er fire universiteter som anvender PURE Portal i dag og det vurderes at tilpasningerne i alt kunne beløbe sig til ca. 500.000,-. Forudsat er, at den centraliserede løsning stiller web services til rådighed, som gør det muligt at foretage universitetsspecifikke søgninger via universitetets hjemmesider.

I det decentrale scenarie, vurderes det, at man vil få behov for at videreudvikle en række elementer i PURE, Orbit og den nationale forskningsdatabase. Det drejer sig bl.a. om:

- Et fælles workflow mellem hhv. PURE, Orbit og den nationale forskningsdatabase i forbindelse med de valideringer som foretages fra centralt hold.
- Implementering af valideringsrutiner i PURE og Orbit, som ikke findes i dag.

¹ Estimateret er baseret på en standardestimeringsmetode med nedbrydning af løsningen i overordnede funktionelle moduler og udviklingen af disse i analyse, design, implementering og test.

Det er vanskeligt at give et estimat af dette, da de afhænger væsentligt af konkrete forhandlinger med leverandørerne af PURE og Orbit, men et gæt kunne være ca. 500.000,-

10.3 IT-Drift decentralt vs. centralt scenarie

Prisen på serverdrift, som er anvendt i nedenstående estimat er et gennemsnitstal fra Gartners benchmark¹ for hvad det koster at drifte en midrange windows server med høj tilgængelighed, server software-support og backup. Den udgør ca. 110.000,-

I et centralt scenarie vil udgifterne til drift bestå af følgende elementer:

Serverdrift af to servere (en indberetningsserver og en statistik og rapporteringsserver)	220.000,-
Applikationssupport og vedligehold	200.000,-
I alt ca.	420.000,-

Hertil kommer eventuelle ekstraudgifter til drift hos de, som i dag udnytter PURE portal, som er baseret på direkte opslag i den lokale PURE database. Forudsætningen for driftsestimatet er dog, at den centrale løsning stiller web services til rådighed, som gør det muligt at foretage søgninger via Universitetets egen hjemmeside i data fra den nationale forskningsdatabase, tilpasset Universitetets behov. Derved er det kun i de tilfælde, hvor man er nødsaget til at etablere en dedikeret web-server løsning til den funktionalitet, som PURE portal dækker nu, at man vil opleve en øgning af IT-driftsudgifterne. Gartner har ikke analyseret dette, men vurderer det som meget usandsynligt, at der vil være nogen Universiteter, som kommer i denne situation.

I et decentralt scenarie vil udgiften til IT-drift kunne estimeres til følgende:

Antal servere centralt	1
Antal servere decentralt	13
Serverdrift centralt	110.000,-
Serverdrift decentralt	1.430.000,-
Applikationssupport	200.000,- ²
I alt ca.	1.740.000,-

Estimatet er baseret på den forudsætning, at Universiteterne leverer IT-drift med priser, opetider, service og support som en gennemsnitlig IT-leverandør.

¹ Gartner driver verdens mest omfattende database over IT-drift og tallet er et gennemsnit af hvad faktiske kunder har betalt for serverdrift, korrigeret for kendte faktorer.

² Baseret på RUC's udgifter til applikationssupport og vedligehold korrigeret for andelen af installationer, som ikke ligesom RUC har PURE portal

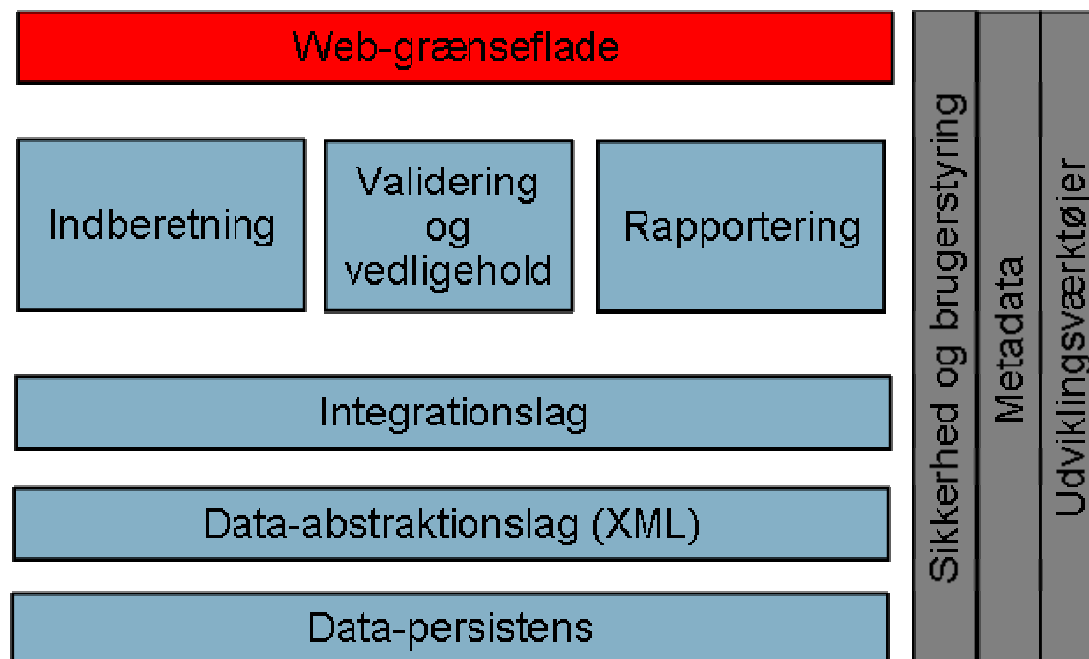
11.0 Krav til en centraliseret løsning

Hvis man vælger scenarie 1 og centraliserer registreringen af forskning til både den bibliometriske forskningsindikator samt anvendelsen til styring og formidling på universiteterne, skal man gennemføre indkøb af en sådan løsning.

Gartner anbefaler en fleksibel og åben arkitektur, som gør det enkelt at justere systemet hen ad vejen.

Særlig er det kritisk for systemets succes, at der udover muligheden for at trække standardrapporter gives en meget fleksibel mulighed for at hente data til både formidlings- og styringsformål via web services.

11.1 Arkitekturtegning til system



11.2 Funktionelle moduler:

I det følgende beskrives kort de moduler, som skal levere direkte funktionalitet til brugere.

Indberetning:

Dette modul skal understøtte forskernes arbejde med at indberette publikationerne. Det vil foregå i form af et komplekst skema med en udbredt brug af feltvalidering. Skemaet har to glimrende forlæg fra PURE og Orbit og kræver derfor et minimum af designaktivitet.

Modulet skal også understøtte at man fra valideringsmodulet kan vise en ny artikel, som stammer fra en søgning i f.eks. ISI. Denne artikel skal herefter checkes og godkendes af forskeren.

Systemet skal kunne notificere brugere via e-mail om artikler eller andet som kræver brugerens handling.

Validering og vedligehold:

Dette modul skal understøtte arbejdet med at validere de indberetninger som er foretaget og generelt vedligeholde databasen.

Modulet skal således understøtte intelligent identifikation af dubletter. Den skal også understøtte søgninger i forhold til et udtræk fra f.eks. ISI.

Det skal være muligt at vedligeholde databasen, dvs. f.eks. tilføje opdateringer fra ISI eller ændringer i den autoritative liste over publiceringskanaler.

Rapportering:

Systemet skal være i stand til at generere standardrapporter til forskellige formål. Det skal anvendes til ledelsesformål på universiteterne på forskellige niveauer.

Det skal være enkelt at eksportere data fra systemet til andre formål. Det kunne være til datawarehouses på universiteterne eller til formidling.

■ ■ ■ ■ Bilag

**Spørgsmål vedrørende denne rapport
bedes adresseret til:**

Peter Suhr
Gartner, Danmark Aps.
E-mail: peter.suhr@gartner.com

Forsknings- og Innovationsstyrelsen Kontaktinformation

Morten Scharff
E-mail: msc@fi.dk