

Teknisk Udvalgs anbefalinger til Styregruppen

1. Baggrund og udgangspunkter for Teknisk Udvalgs anbefalinger

Teknisk Udvalg blev nedsat i juni 2007 og havde som overordnet opgave at bistå Forsknings- og Innovationsstyrelsen med at udvikle en tværgående forskningsdatabase, der indeholder valide og sammenlignelige data over publicering blandt danske forskere fra 2008.

Udvalget blev nedsat i juni 2007 og afsluttede sit hidtidige arbejde ved årsskiftet 2007/2008 med nærværende forslag til styregruppen om den fremtidige organisering og drift af den tværgående forskningsdatabase. Herefter skal Teknisk Udvalg følge etableringen af den tværgående forskningsdatabase og i løbet af 2008 udarbejde et forslag til en permanent organisering af databasen.

Udvalget har taget afsæt i de principper, som blev fastlagt i Styregruppens principppapir fra maj 2007 og har løbende indarbejdet de retningslinier og beslutninger, som Styregruppen og Fagligt Udvalg har taget.

Til brug for sit arbejde har Teknisk Udvalg fået udarbejdet to større analyser. Det drejer sig om følgende undersøgelser/analyser:

- **Grundpapir.** Grundpapiret giver et overblik over det samlede systemlandskab på forskningsregistreringsområdet. For de enkelte delområder i systemlandskabet beskriver grundpapiret status for den nuværende situation i Danmark, udviklingsbehov og udfordringer set i lyset af en kommende forskningsdatabase. Herudover belyser grundpapiret behovet for yderligere viden. Grundpapiret blev udarbejdet af ansatte ved DTV og KU/Life og blev afsluttet i oktober 2007.
- **Konsulentundersøgelse.** På baggrund af ovenstående grundpapir igangsatte Teknisk Udvalg en ekstern konsulentundersøgelse, som havde til formål at udarbejde et forslag til udvikling af en tværgående forskningsdatabase, herunder et system til indsamling, registrering og validering af publikationsdata. Konsulentundersøgelsen skulle endvidere belyse, hvordan arbejdet med databasen kunne organiseres, herunder udarbejde et forslag til grænsefladerne og samarbejdet mellem de lokale indberetningssystemer og en tværgående forskningsdatabase. Konsulentundersøgelsen blev udført af Gartner Group og blev afsluttet december 2007.

11. marts 2008

Forsknings- og Innovationsstyrelsen

Bredgade 40
1260 København K
Telefon 3544 6200
Telefax 3544 6201
E-post fi@fi.dk
Netsted www.fi.dk

Sagsbehandler
Morten Scharff
Telefon 7226 5530
Telefax 3544 6201
E-post msc@fi.dk

Sagsnr. 08-031726
Dok nr. 486218
Side 1/8

2. Forskningsregistreringslandskabet

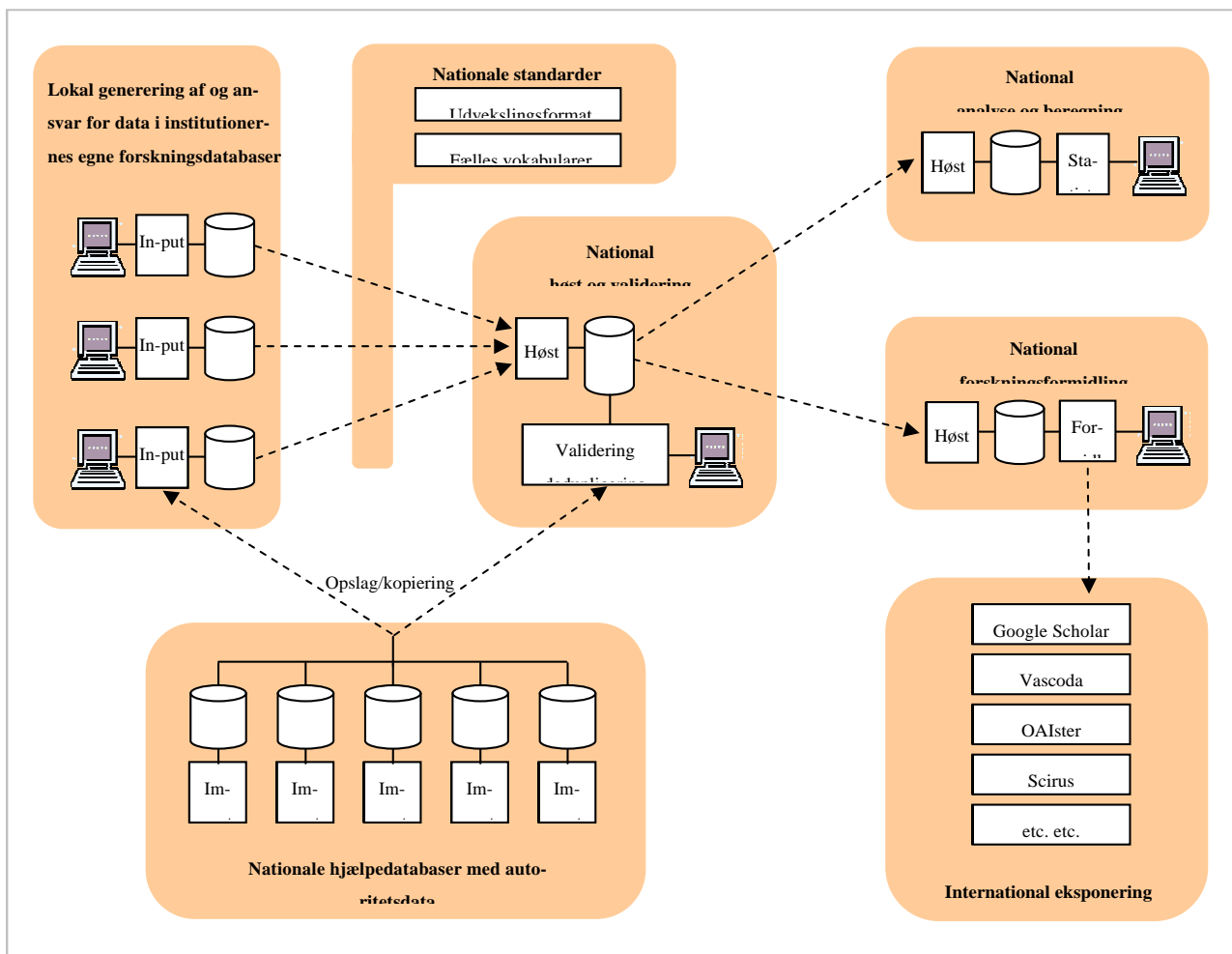
I Danmark er der en lang tradition for at registrere forskernes publikationer lokalt på universiteterne. Der har været forskellige formål og forskellige parter involveret i denne registrering. På den ene side findes forskningsinstitutionerne og forskningsbibliotekerne, som har registreret forskernes publikationer med det formål at formidle viden om dansk forskning og danske forskningspublikationer. På den anden side findes "forskningsadministrationerne", som typisk har behov for information om forskernes output til ledelsesinformation. I dag har de fleste lokale forskningsdatabaser en høj dækningsgrad og anvendes til en bred vifte af formål på institutionerne - lige fra forskningsformidling til ledelsesinformation. Den tekniske og den proceduremæssige viden i dette arbejde har udviklet sig løbende og er akkumuleret på hver sin måde i de to systemer Orbit (DTU) og PURE (AU, AAU, CBS, KU, SDU, RUC). Kendetegnende for begge systemer er, at de understøtter høstning via et åbent udvekslingsformat for metadata.

Diagrammet nedenfor illustrerer skematisk det fremtidige forskningsregistreringssystem i Danmark og er opdelt i en række funktionsområder med hver sin problematik og faglighed.

Forsknings- og Innovationsstyrelsen

Side 2/8

Figur 1: Systemlandskabet



- A. **Universiteternes lokale forskningsdatabaser.** Disse er basis for datagenereringen og leverer datagrundlaget for den tværgående forskningsdatabase. I de senere år har alle universiteter installeret moderne systemer på området - PURE eller Orbit (med undtagelse af ITU, som har deres eget system). Med en ny tjeneste med nationale hjælpe-databaser vil det lokale inddateringsarbejde kunne lettes og datakvaliteten øges.
- B. **Nationale udvekslingsformater og vokabularer for forskningsmetadata** sikrer en præcis og korrekt udnyttelse/repræsentation af universiteternes (og andre forskningsinstitutioners) lokale data i det videre nationale system.
- C. **National indsamling (høst) samt teknisk/bibliografisk validering af forskningsdata.** Indsamler løbende og automatisk de lokale data fra universiteternes lokale forskningsdatabaser samt sikrer deres validitet i forhold til udvekslingsformat og vokabularer. Med en ny tjeneste med nationale hjælpe-databaser udvides den automatiske validering til også at omfatte publiceringskanaler. I tillæg til validering vil der antagelig blive behov for central deduplikering eller sammenkædning af metadata for flere registreringer af den samme publikation - typisk når forfattere fra flere institutioner har arbejdet sammen. Dette kan understøttes af software, men vil givet også kræve menneskelig indsats for at opnå den tilstrækkelige præcision.
- D. **Nationale hjælpe-databaser med autoritetsdata.** Disse vil omfatte metadata om publikationskanaler som f.eks. tidsskrifter og bogforlag samt persondata.
- E. **National beregning af bibliometrisk forskningsindikator.** Data udtrækkes af den validerede nationale indhøstning til forskningsadministrative formål i ministerieregi.
- F. **National forskningsformidling med internationale aspekter.** Data udtrækkes af den validerede nationale indhøstning til støtte for formålet national og international eksponering og formidling af danske forskningsresultater. Den nationale portal og søgemaskine indgår i samarbejde med Google Scholar og andre internationale platforme, der kan styrke eksponering og formidling.

2.1 Udvalgets vurderinger af systemlandskabet

Det er udvalgets vurdering, at det eksisterende forskningsregistreringssystem, som er opbygget på såvel de danske universiteter som i regi af DEFF, kan danne udgangspunkt for etableringen af den tværgående forskningsdatabase. Det betyder, at der ikke er behov for at opbygge helt nye systemer, men at der kan tages afsæt i de systemer og praksisser, som allerede er etableret.

I Danmark har man siden 2006 anvendt udvekslingsformatet DDF-MXD (Danish Research Database - Metadata Exchange Format for Documents) til overførsel af metadata fra de lokale forskningsdatabaser til den nationale forskningsdatabase. Formatet er udviklet i DEFF-regi i årene 2004-05 i et samarbejde mellem universiteter og den nationale forskningsdatabase. Formatet er fuldt tilpasset og understøttet af såvel PURE- som Orbit-systemerne.

Det er således udvalgets vurdering, at formatet er tilstrækkeligt udviklet til at kunne anvendes af den tværgående forskningsdatabase. På en række områder er der dog behov for mindre tilpasninger af formatet.

I regi af DEFF er der etableret et system til høstning af de "lokale" forskningsregistreringer, som på mange områder dækker behovene for den tværgående forskningsdatabase. Det er udvalgets vurdering, at der ikke er problemer med høstning eller samkøring af data på tværs af de to systemer. Den tværgående forskningsdatabase kan i 2008 derfor tage afsæt i de principper, som i dag benyttes til høstning af data til Den Danske Forskningsdatabase (DDF). Forskerne vil således ikke blive pålagt ekstra registreringsopgaver og vil kun i meget begrænset omfang opleve ændringer i den måde, de registrerer forskningspublikationer på. De skal som hidtil registrere deres publikationsaktiviteter i de lokale registreringssystemer, hvorefter data automatisk høstes eller samles i den tværgående forskningsdatabase.

Det er udvalgets vurdering, at der i dag sker en validering og kvalitetssikring af data, såvel decentralt som centralt, og at de igangværende aktiviteter i et vist omfang dækker behovet for kvalitetssikring af de registrerede forskningspublikationer. Det er dog samtidig vurderingen, at der på en række områder vil være behov for at øge kvaliteten af de indsamlede data (se efterfølgende afsnit).

Forsknings- og
Innovationsstyrelsen

Side 4/8

Det er udvalgets vurdering, at der i konceptet omkring Den Danske Forskningsdatabase er grundlag for at opbygge den tværgående forskningsdatabase, både hvad angår høstning af data (input), selve databasens opbygning, skaleringsmuligheder og funktionalitet samt mulighederne for igen at trække data ud af basen til brug for beregningen af forskningsindikatoren (output).

3. Udfordringer og udviklingsbehov for den tværgående forskningsdatabase

Selv om de eksisterende systemer på mange områder er velfungerende og har en høj dækningsgrad, så er der på en række områder behov for at tilpasse og udvikle systemerne, så de matcher behovene i en kommende forskningsdatabase. Teknisk Udvalg har primært identificeret to typer af udfordringer eller udviklingsbehov. For det første er der et behov for at tilpasse den eksisterende registreringspraksis til de principper, som Styregruppen og Fagligt Udvalg har fastlagt vedr. publikationsformer. Det drejer sig typisk om en række mindre tilpasninger af de eksisterende systemer. For det andet er der på en række områder behov for at styrke kvaliteten af de indsamlede data. Der er her behov for at igangsætte en række aktiviteter, der tilsammen kan højne datakvaliteten af forskningsdatabasen.

Tilpasning og konsekvensrettelser

På en række områder er der behov for at tilpasse den eksisterende registreringspraksis til de principper, der er beskrevet i Styregruppens princippapir fra maj 2007, eller som efterfølgende er blevet fastlagt af Fagligt Udvalg/Styregruppen. Det drejer sig om følgende områder:

- *Tilpasning af registrerings- og udvekslingsformat.* Der er behov for at tilpasse det eksisterende registrerings- og udvekslingsformat, så det matcher Styregruppens beslutning vedr. kategorisering af publikationsformer. Teknisk Udvalg har på den baggrund udarbejdet forslag til kategorisering af publikationsformer. **Udvalget anbefaler, at notatet danner baggrund for den videre tilpasning af DDF-MXD-formatet.**
- *Hovedområderegistrering.* For at matche behovene i kvalitetsfinansieringsmodellen er det afgørende, at publikationsdata i den tværgående

forskningsdatabase entydigt kan kategoriseres inden for de forskellige hovedområder. Det betyder, at der vil være behov for, at registreringen af publikationer i PURE og Orbit også skal indeholde oplysninger om hvilket hovedområde, publiceringen skal kategoriseres under. **Udvalget anbefaler, at hovedområderegistreringen sker på baggrund af forskernes ansættelsesforhold**, således at forskere, der eksempelvis er ansat på det humanistiske fakultet, får registreret sine publikationer som humanistisk forskning.

Styrkelse af datakvalitet

Det er helt afgørende, at kvaliteten af data i den fælles forskningsdatabase er høj, således at der ikke kan sættes spørgsmålstejn ved forskningsdatabasens validitet. På den baggrund er det udvalgets vurdering, at der er behov for at gennemføre en række tiltag, som kan sikre en højere datakvalitet. Tiltagene retter sig såvel mod de lokale indberetningssystemer, som mod den centrale validering og kvalitetskontrol. Udvalget har identificeret tre primære indsatsområder:

- *Hjælpedatabaser og autoritetsdata.* Teknisk Udvalg vurderer, at der til støtte for den lokale forskningsregistrering på universiteterne og andre forskningsinstitutioner skal opbygges hjælpedatabaser, som skal anvendes af forskningsinstitutionerne, når de registrerer forskernes publikationer i lokale forskningsregistreringssystemer. Vurderingen er, at hjælpedatabaserne/autoritetsdata vil være en vigtig ressource til at styrke kvaliteten af de indberettede data. Der er primært to områder, hvor der er behov for hjælpedatabaser med autoritetsdata:
 - *Hjælpedatabase med publikationskanaler.* Det vurderes, at der på baggrund af faggruppernes indstillinger, skal etableres en hjælpedatabase med autoritetsdata over godkendte publikationskanaler (forlag og tidsskrifter). Hjælpedatabasen skal kobles til de lokale systemer, så ajourførte data altid er tilgængelige i den lokale registreringsproces. Det er vurderingen, at et sådant system vil mindske antallet af fejlindberetninger. **Teknisk Udvalg anbefaler, at etablering og drift af hjælpedatabase overlades til en operatør.**
 - *Forskerdatabase.* Det vurderes, at der skal etableres en hjælpedatabase med oplysninger om forskere (navn, stilling, og ansættelsessted), der er ansat på de danske forskningsinstitutioner. Hjælpedatabasen gøres tilgængelig i den lokale indberetningsproces, således at det i forbindelse med registrering af publikationsdata bliver muligt at udpege medforfattere på tværs af forskningsinstitutionerne. Det vil lette indberetningen ved flerforfatterskaber. **Teknisk Udvalg anbefaler, at etablering, drift og vedligehold af hjælpedatabase overlades til en operatør. I første omgang anbefales, at der etableres en forskerdatabase, som samler og systematiserer oplysninger fra de forskellige universiteter.** Det kan på sigt overvejes, om databasen skal udvikles således, at den enkelte forsker forsynes med et unikt ID, der gør det muligt at identificere forskere med flere ansættelsessteder.
- *Central validering/kvalitetskontrol.* Der stilles store krav til kvaliteten af data i forskningsdatabasen. Selv om universiteterne allerede i dag gør en

stor indsats for at mindske fejlindberetninger etc., er det udvalgets vurdering, at der vil være behov for at etablere en "central" validering/kvalitetskontrol. **Udvalget anbefaler, at én eller flere personer (ansat uafhængigt af universiteterne) får til opgave at gennemføre en stikprøvekontrol af de indberettede data, herunder at sikre en ensartet indberetningspraksis på tværs af forskningsinstitutionerne.** På sigt skal det undersøges om en central validering/kvalitetskontrol skal udbygges, så der sker en automatisk validering pba. hjælpedatabaser med information om publikationsdata (eks. i forbindelse med udvikling af det "databrøndsprojekt" som er under udvikling i DEFF-regi).

- *"Deduplikeringsproblemet"*. Det er udvalgets vurdering, at der er behov for at mindske antallet af "dobbelregistreringer". I dag findes der ingen deduplikeringsmekanisme i DDF. Det er et problem, fordi der ofte publiceres publikationer med forfattere fra flere forskellige danske universiteter. Det betyder, at der er publikationer, som bliver registreret to eller flere gange. En del af *"deduplikeringsproblemet"* må forventes at blive løst i takt med, at der etableres en forskerdatabase. Det er dog udvalgets vurdering, at der stadig vil være behov for at etablere en "central" validering/kvalitetskontrol, som skal identificere dobbeltregistreringer ved flerforfatterskaber. **Udvalget anbefaler, at én eller flere personer (ansat uafhængigt af universiteterne) får til opgave at identificere "dobbelregistreringer" samt koble indberetninger på tværs af forskningsinstitutionerne i forbindelse med flerforfatterskaber.**
- *Teknisk Arbejdsgruppe*. Teknisk Udvalg anbefaler, at FI nedsætter en teknisk arbejdsgruppe med repræsentanter fra Atira, DTIC, universiteterne og DTU, som skal udmønte og implemetere Teknisk Udvalgs anbefalinger. Det foreslås, at Teknisk Arbejdsgruppe refererer til FI og ledes af en teknisk projektleder ansat af FI.

4. Udvalgets anbefalinger for etablering af den tværgående forskningsdatabase

Forskningsdatabase skal indeholde registreringer af publikationsaktiviteter i 2008. Det betyder, at der er meget kort tid til at etablere en funktionsdygtig database. På den baggrund anbefaler Teknisk Udvalg, at udviklingen af forskningsdatabase tager udgangspunkt i de eksisterende systemer for forskningsregistrering, høstning og central lagring i DDF. Det vurderes som realistisk at kunne tilpasse og udvide disse systemer til, at DDF vil kunne fungere som den tværgående forskningsdatabase og datagrundlag for forskningsindikatoren.

Teknisk Udvalg anbefaler, at de eksisterende forskningsregistreringssystemer i 2008 fortsætter som hidtil med at registrere forskernes publikationsaktiviteter (dog med den række af mindre justeringer og tilpasninger, som er beskrevet i det ovenstående afsnit). Det betyder, at det også fremover er forskningsinstitutionernes ansvar at indsamle og gøre data om forskernes publikationsaktiviteter tilgængelige.

Det betyder, at Den Danske Forskningsdatabase (DDF) i regi af Danmarks Tekniske Informationscenter (DTIC) som hidtil fortsætter med at "høste" data fra de

lokale forskningsregistreringssystemer. Den tværgående forskningsdatabase bygger således på de data fra 2008, som ”høstes” af DDF. Som udgangspunkt ændres der ikke ved de principper eller metoder, som benyttes til at høste data. På enkelte områder vil der dog være behov for at udvide eller supplere de eksisterende aktiviteter, som gennemføres af DTIC. Det drejer sig bl.a. om at udarbejde udtræk fra forskningsdatabasen til brug for validering af data samt beregning af forskningsindikator. **Udvalget anbefaler i den sammenhæng at DEFF, DTIC og FI indgår aftale om udviklingen af Den Danske Forskningsdatabase for 2008, og at udvalget efterfølgende orienteres om aftalens indhold.**

Der er som tidligere nævnt behov for at etablere en hjælpedatabase med publikationsdata samt en forskerdatabase. Udvalget anbefaler, at FI udpeger en eller flere operatører til at udvikle og drive de to hjælpedatabaser, herunder også at gøre hjælpedatabaserne tilgængelige for universiteterne og andre forskningsinstitutioner (disse indarbejder efterfølgende listerne i registreringssystemer).

For at styrke datakvaliteten i forskningsdatabasen anbefaler Teknisk Udvalg, at der uafhængigt af universiteterne etableres en validering og kvalitetskontrol af de indsamlede publikationsdata (således at der ikke kan stilles spørgsmål ved uafhængigheden af den centrale validering og kvalitetskontrol). Det vurderes, at der i 2008 primært bør være fokus på to områder: A) identifikation af dobbeltregistreringer ved flerforfatterskaber og B) test af datakvalitet i form af stikprøvekontrol blandt indsamlede data.

Det skal understreges, at selv om der ifølge udvalgets vurdering ikke kræves store tekniske indgreb i de eksisterende systemer, så er der tale om distribuerede systemer med en stor organisatorisk kompleksitet og mange mennesker involveret i registreringen af forskningsdata. Det betyder, at selv begrænsede tekniske ændringer tager tid at specificere, implementere og teste på en måde, der kan sikre datakvaliteten hele vejen gennem systemerne. Det er derfor en betingelse for succes, at processen sættes i gang meget hurtigt, så de nødvendige ændringer i indberetningsgrænsefladerne kan være til rådighed for universiteterne ved størstedelen af indberetningerne for 2008.

5. Udvalgets anbefalinger på lang sigt

Rapporten *En national forskningsdatabase* fra Gartner Group (21-12-2007) beskriver flere mulige scenarier for den tekniske organisation af forskningsdatabasen på lang sigt. Det er Gartners vurdering, at der vil kunne være en driftsøkonomisk gevinst ved at erstatte det distribuerede koncept med et enkelt centralt system, der integrerer forskningsregistreringen direkte i den tværgående forskningsdatabase. Gartner har også påpeget muligheden for at man, hvis det distribuerede koncept bevares, kunne lade driften af den tværgående forskningsdatabase gå i udbud.

Teknisk Udvalg har vurderet, at der, udover det i snæver forstand driftsøkonomiske, indgår en række andre faktorer i vurderingen af mulige fremtidsscenarier, især forhold omkring universiteternes ejerskab og muligheder for lokale anvendelser af de indberettede data.

Det er derfor udvalgets anbefaling, at der, når etableringen af registrering og høstning for 2008 er godt i gang, iværksættes en undersøgelse af den fremtidige organisering af forskningsregistreringen, som, ud over driftsøkonomi, tager højde for disse forhold.

6. Den videre proces

Det er vurderingen, at der også efter januar 2008 vil være behov for et teknisk udvalg til at følge udviklingen og etableringen af den tværgående forskningsdatabase. Det anbefales derfor, at Teknisk Udvalg fortsætter sit arbejde i 2008. Det foreslås endvidere, at der udarbejdes et nyt kommissorium for udvalget, som bl.a. indeholder følgende målsætninger for udvalgets arbejde:

- Det anbefales, at Teknisk udvalg fremover følger udviklingen og opbygningen af forskningsdatabasen. Udvalget følger indsamlingen og høstningen af data for forskningspublicering i 2008. Der er bl.a. fokus på at følge den centrale validering og kvalitetskontrol, som gennemføres i regi af Den Danske Forskningsdatabase. Udvalget vurderer løbende kvaliteten af de indsamlede publikationsdata. Teknisk Udvalg skal endvidere følge opbygningen af validerings- og kvalitetskontrollen, samt komme med forslag til hvordan datakvaliteten af de indsamlede data kan højnes.
- I forbindelse med konkrete uenigheder eller tvivlsspørgsmål omkring fortolkning af de indsamlede data rådgiver udvalget FI og Den Danske Forskningsdatabase om, hvordan registreringen af forskningspublikationer bør gennemføres.
- Teknisk Udvalg undersøger nye muligheder for organiseringen af den tværgående forskningsdatabase efter 2008 og udarbejder et samlet forslag til den fremtidige organisering af databasen i løbet af efteråret 2008.
- Udvalget sekretariatsbetjenes af Forsknings- og Innovationsstyrelsen, der også afholder udgifter i forbindelse med møder og øvrig sekretariatsbi-stand.

FI nedsætter en teknisk arbejdsgruppe til udmøntning og implementering af Teknisk Udvalgs anbefalinger.