
Kravspecifikation

SwePub

www.swepub.se

UTKAST FÖR REMISS
2008-04-02

Innehållsförteckning

INNEHÅLLSFÖRTECKNING	2
1. INTRODUKTION	4
1.1. BAKGRUND	4
1.2. SYFTE OCH MÅL	6
1.3. DELTAGARE I PROJEKTET	6
2. KONCEPTUELL ÖVERSIKT	8
3. ANVÄNDARGRUPPER OCH RÄTTIGHETER	9
3.1. PUBLIKA ANVÄNDARE AV SÖKGRÄNSSNITTET	9
3.2. ANVÄNDARE AV UTVÄRDERINGSGRÄNSSNITTET	9
3.3. LOKALA ADMINISTRATÖRER	9
3.4. SUPERADMINS	9
4. DATAMODELL	10
5. DATAINHÄMTNING	11
5.1. LOKALA DATABASER FÖR PUBLICERING	11
5.1.1. <i>Definition</i>	11
5.1.2. <i>Protokoll för dataöverföring</i>	11
5.1.3. <i>Krav på lokala system</i>	11
5.2. KRAV PÅ CENTRALT SYSTEM	12
6. DATAFÖRBÄTTRING	13
6.1. VERIFIERING	13
6.2. DEDUPLICERING.....	13
6.3. BERIKNING FRÅN ANDRA DATABASER	14
6.4. KOPPLING TILL PERSON.....	14

7. SÖKGRÄNSSNITT	15
7.1. ÖVERSIKT OCH INNEHÅLL	15
7.2. RELATION TILL BEFINTLIGA SYSTEM.....	15
7.3. PRINCIPIELLA FUNKTIONER.....	16
7.4. ADMINISTRATION AV PERSONDATA	17
8. UTVÄRDERINGSGRÄNSSNITT	18
8.1. ÖVERSIKT OCH INNEHÅLL	18
8.2. RELATION TILL ÖVRIGA SYSTEM.....	18
8.3. PRINCIPIELLA FUNKTIONER.....	19
8.3.1. <i>Ingående parametrar i analysverktyget.....</i>	<i>19</i>
8.3.2. <i>Förslag till ytterligare parametrar utifrån aktörsintervjuerna</i>	<i>20</i>
9. ICKE-GRAFISKA EXPORTGRÄNSSNITT	21
9.1. OAI-PMH.....	21
9.2. EXCEL (CSV)	21
9.3. ANDRA FORMAT.....	22
9.4. GOOGLE (OCH ANDRA SÖK-TJÄNSTER).....	22
10. PROJEKTBESKRIVNING OCH TIDPLAN	23
11. STYRORGANISATION.....	24
11.1. UNDER PROJEKTET	24
11.2. I DRIFTSLÄGET	24
11.3. FRÅN PROJEKT TILL KONTINUERLIG DRIFT	25
BILAGA	27

1. Introduktion

1.1. Bakgrund

Detta projekt - *Samlad ingång till och redovisning av svensk vetenskaplig publicering* - är resultatet av en projektutlysning¹ som gjordes i mars 2007 inom programmet OpenAccess.se² vid Kungl. biblioteket (KB). Projektet kommer att utveckla

- en publik söktjänst inom ramen för Libris webbsök
- en nationell databas som kan utnyttjas av andra tjänster för redovisning och analys av den svenska vetenskapliga publiceringen
- ett webbaserat analysverktyg för generering av rapporter baserade på den nationella databasen

Dessa tjänster har det gemensamma arbetsnamnet *SwePub*.

Övergripande motiv för KB:s projektutlysning var att

- ge ökad synlighet och tillgänglighet för svensk vetenskaplig publicering
- understödja en utveckling där en växande del av den svenska vetenskapliga publiceringen är fritt tillgänglig i fulltext på Internet
- ge ett samlat och jämförbart underlag för redovisning av svensk vetenskaplig publicering

¹ Se: <http://www.kb.se/om/projekt/openaccess/nyheter/2007/Utlysning-till-projekt/>

² Se: <http://www.kb.se/om/projekt/openaccess/>

- ge incitament för att lokala publiceringsdatabaser ska bli heltäckande och följa överenskomna standarder³

Tjänstens potentiella användare är

- forskare, lärare, studenter
- andra beslutsfattare inom högskola och forskning
- användare utanför högskolan som behöver en samlad ingång till den svenska vetenskapliga publiceringen

KB:s förutsättningar för arbetet såsom de angavs i utlysningen är att

- arbetet ska utgå från de rekommendationer för ett nationellt format för publiceringsdatabaser (lokala register över akademisk publicering)⁴ som utarbetats inom *SVEP-projektet*⁵ och godkänts av SUHF
- arbetet ska bygga vidare på resultaten från projektet *Söktjänst för svenska vetenskapliga publikationer*⁶
- söktjänsten ska utvecklas och tas i permanent drift av avdelningen för LIBRIS som en söktjänst inom LIBRIS nationella system och metadata ska kunna ingå i LIBRIS webbsök. LIBRIS har huvudansvar för att:
 - skapa en databas; hämtning av data, indexering och uppbyggnad m.m.
 - skapa en söktjänst; bygga logik, funktionalitet, visningsformat mm.
- detta arbete ska ske i nära samverkan med ansvariga för det utlysta projektet
- data som insamlas för tjänsten ska fritt kunna användas av andra parter för olika slag av publika tjänster

³ Enligt rapporten *Kartläggning av högskolans elektroniska publicering och rekommendationer för nationell samverkan* har 25 av de 42 tillfrågade lärosätena svarat att man driver en tjänst för publiceringsredovisning, se: <http://www.suhf.se/BinaryLoader.aspx?OwnerID=51bfc2f3-8841-4d55-b08e-e54dfc8e8014&OwnerType=2&ModuleID=a8429cd5-a9d9-4e2b-a1d8-4a7cf26cd7a3&PropertyCollectionName=Content&PropertyName=Files&ValueIndex=23>

⁴ Se: http://epc.uu.se/files/publ1_2.pdf

⁵ Se: <http://www.kb.se/om/projekt/openaccess/svep/>

⁶ Se: http://epc.uu.se/files/ssvp-final1_1.pdf

- insamling av metadata förutsätts i första hand göras med stöd av OAI-PMH (Open Archives Initiative - Protocol for Metadata Harvesting)

1.2. Syfte och mål

Projektet ska utveckla en tjänst som:

- hämtar samtliga poster för vetenskapliga publikationer från lärosätenas publiceringsdatabaser
- gör dessa poster tillgängliga för sökning från slutanvändare och för hämtning till andra tjänster
- ger möjlighet och verktyg för att utnyttja metadata som underlag för redovisning och analys av den svenska vetenskapliga publiceringen
- fungerar som en publikt tillgänglig söktjänst inom ramen för Libris nationella system

Projektet ska särskilt

- se över kraven på metadata och utforma en kravspecifikation för tjänsten
- kommunicera med alla högskolor angående utformning och deltagande i tjänsten
- utforma informationsmaterial och marknadsföring av tjänsten
- utveckla verktyg för att utnyttja metadata från tjänsten som underlag för redovisning och analys av den svenska vetenskapliga publiceringen.

1.3 Deltagare i projektet

- Uppsala universitetsbibliotek
- Lunds universitetsbibliotek
- Göteborgs universitetsbibliotek
- Linköpings universitetsbibliotek
- Avedas AG

samt

- Avdelningen för Libris, Kungl. biblioteket

För att inhämta information om vad som är viktigt att ha med i kravspecifikationen har projektet i förberedelsefasen även träffat ett antal externa aktörer:

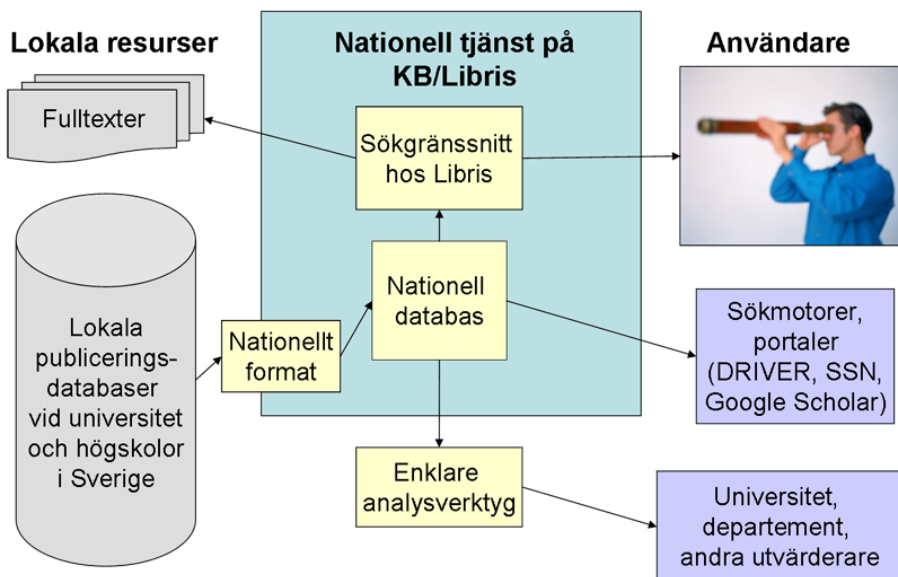
- Fridasystemet i Norge
- Forskningsdatabasen i Danmark
- Sweden Science Net
- Intervju med RUT2-utredningens⁷ huvudsekreterare Kjell Blücker (forskningssekreterare på Riksbankens jubileumsfond)
- SUHF-Bibliometri/Metrics-gruppen
- SUHF-Publicering
- SUHF, generalsekreteraren
- VR, analysenheten
- VR, Forskning.se
- Universitetsledningarna vid de deltagande universiteten
- Intervjuer med forskare
- Intervjuer med informationsansvariga vid Linköpings universitet
- Intervjuer på de deltagande biblioteken

⁷ Resurser för kvalitet, SOU 2007:81, <http://www.regeringen.se/sb/d/8439/a/91339>

2. Konceptuell översikt

De två olika applikationerna som utvecklas i projektet - den publika söktjänsten och analysverktyget - kommer att hämta data från den gemensamma databasen som hämtats från de lokala systemen. Det nationella formatet anpassar de lokala posterna till en enhetlig modell. Via söktjänsten skall man i så stor utsträckning som möjligt kunna nå även fulltexter, antingen i lokala öppna arkiv eller i andra databaser. Den nationella databasen kommer också att kunna leverera metadataposter till andra söktjänster via OAI-PMH.

Söktjänst för svenska vetenskapliga publikationer



3. Användargrupper och rättigheter

3.1. Publika användare av sökgränssnittet

Sökning i det publika gränssnittet ska inte kräva någon autentisering och vara fritt tillgängligt.

3.2. Användare av utvärderingsgränssnittet

Användning av utvärderingsgränssnittet ska inte kräva någon autentisering och vara fritt tillgängligt.

3.3. Lokala administratörer

Lokala databasinstitutioner ska kunna kontrollera höstningsresultatet och inställningar. För detta krävs inloggning och registrering. Separat användardatabas upprättas.

3.4. SuperAdmins

Administratörer vid Libris ska kunna logga in separat.

4. Datamodell

En uppdatering av metadatamodellen i *Nationellt format för publiceringsdatabaser* har gjorts och bifogas här som särskild bilaga. Poster skall levereras till den nationella databasen i metadataformatet *MODS (Metadata Object Description Schema)*⁸ enligt en särskild profil – SwePub MODS.

⁸ Se: <http://www.loc.gov/standards/mods/>

5. Datinhämtning

SwePub bygger på att innehållet i databasen hämtas från lokala databaser. För att detta skall kunna fungera på ett rimligt sätt behöver inhämtningsmetod och postformat standardiseras.

5.1. Lokala databaser för publicering

5.1.1. Definition

Databaser som regelbundet registrerar svenska lärosätens produktion av publikationer (publiceringsdatabaser).

5.1.2. Protokoll för dataöverföring

De lokala databaserna måste stödja OAI-PMH 2.0⁹.

5.1.3. Krav på lokala system

Lokala system skall kunna tillhandahålla ett OAI-PMH set där posterna är formaterade efter SwePubs MODS-specifikation (eller skapa SwePub MODS genom transformering i det centrala administrationsgränssnittet). Poster i den lokala databasen som redovisar publikationer som publicerats när författaren inte var knuten till lärosätet skall inte ingå.

⁹ Se: <http://www.openarchives.org/OAI/openarchivesprotocol.html>

Varje lärosäte skall leverera obligatoriska element enligt SwePub MODS, men det är valfritt att skicka med icke obligatoriska element i formatspecifikationen.

Teckenkodningen ska vara UTF-8.

5.2. Krav på centralt system

Administrativt gränssnitt för kontroll av hämtningar omfattar:

- Funktion för att identifiera och rapportera fel
- Kvittens / Information / Statistik om hämtningar som gjorts
- Konfigurera inställningar (serveradress, set som hämtas, hämtningsfrekvens och datum)
- Transformering (editeringsbart stylesheet som transformerar lokalt metadataformat till SwePub MODS)
- Det centrala systemet skall stödja delete-funktionaliteten, dvs att enbart förändrade eller nya poster hämtas, i OAI-PMH.

6. Dataförbättring

När höstningen gjorts innehåller data en del fel, många som inte uppkommer förrän data aggregerats. Detta gäller t.ex. förekomst av dubletter. För att informationen ska vara pålitlig för diverse tillämpningar, så behövs en del kvalitetshöjande arbete göras.

6.1. Verifiering

Inhästad data ska verifieras innan den tillförs den nationella databasen. Detta görs genom att Libris kontrollerar att höstningen var komplett samt genom att kontrollera att obligatoriska fält finns med i samtliga poster.

6.2. Deduplicering

När forskare från olika svenska universitet samarbetar, finns det risk för att dubletter uppstår i Swepub-databasen. Identifikation av dubletter och deduplicering kräver utveckling av nya algoritmer. Om det är tekniskt möjligt ska publikationer med identiska externa id (från Thomson, Scopus eller Pubmed) kopplas samman på ett sådant sätt att den presenteras och räknas som en post i träfflistor (hos Libris) och i utvärderingar i analysverktyget. Libris ska också undersöka om en kombination av identiska fält mellan två poster ska leda till deduplicering.

6.3. Berikning från andra databaser

Flera aktörer förespråkar att databasen berikas med data från externa källor. Kopplingar till externa databaser ska om möjligt göras lokalt och det rekommenderas att externa id förs över till Swepub. Kopplingar mellan böcker och bokkapitel till Libris stora databas ska göras så långt som möjligt så att korssökning kan göras på ett tilltalande sätt i sökgränssnittet.

6.4. Koppling till person

Identiteter från lokal verifiering ska överföras till Swepub och användas vid sökning och analys.

7. Sökgränssnitt

Huvudsaklig målgrupp för det externa sökgränssnittet förväntas vara forskare och övriga externa intressenter, medan de som behöver information för olika former av verksamhetsanalys, statistik, bibliometri etc. förväntas använda den speciella analys- och statistikfunktionen alternativt någon form av exportgränssnitt.

7.1. Översikt och innehåll

Som grund för arbetet med det externa sökgränssnittet har använts resultat från intervjuer, som genomförts med olika aktörer.

Det är viktigt att sökgränssnittet blir en bra kompromiss mellan å ena sidan behovet av detaljerad information, och å andra sidan den enkelhet som krävs för att t.ex. forskare ska se mödan värt att använda systemet.

7.2. Relation till befintliga system

En utgångspunkt i det inledande arbetet har varit att använda Libris Webbsök för det nya sökgränssnittet i SwePub. Detta inkluderar också den grafiska utformningen, vilket underlättar för användaren som då "känner igen sig" och inte behöver tänka om mellan olika system.

Principiellt kan man välja mellan att mer eller mindre integrera SwePub och Libris alternativt att "kopiera" Librisformen till en ny SwePub-applikation, som då utgör ett " eget gränssnitt". Huvudinriktningen i diskussionerna har varit att sträva mot en integrering, men detta måste utredas mer med hjälp av

användarundersökningar. Analyser kommer att utföras under hela utvecklingsprocessen och det alternativ som slutligen väljs skall svara mot användarens behov (användarcentrerad design).

Beroende på vilken principiell uppbyggnad som väljs utifrån användarundersökningarna får sedan en lämplig anpassning göras till de behov som i övrigt växer fram i SwePub-arbetet.

Det är också nödvändigt att beakta vilken form av data som används och utföra tester för att se hur data skall hanteras, och vad som är möjligt när det gäller dataöverföring mellan Libris MARC och SwePub MODS.

7.3. Principiella funktioner

Användaren skall kunna välja mellan enkel och avancerad (utökad) sökning. Enkel sökning utformas som en fritextsökning, vilken indexerar flertalet fält.

Avancerad sökning utformas som en sökavgränsning alternativt i form av ett specifikt utökat gränssnitt (beroende på vald modell enligt 7.2 ovan).

Utökad sökning (specifikt gränssnitt) bör innehålla följande funktioner: (skall kunna kombineras vid sökning)

- Årtal/tidsperiod
- Namn (författare/redaktör)
- Lärosäte
- Fakultet
- Institution
- Titel, ord i titel
- Ämne/ämnesområde (vetenskapsområde)
- Publikationstyp/rapportklass
- ISBN/ISSN
- Tidskrift
- Språk

Övriga önskemål:

- Koppling till fulltext med hjälp av DOI, SFX etc. liksom möjlighet att i sökträff se om fulltext är tillgänglig.
- RSS-funktion med möjlighet att se vad som löpande publiceras i ett visst ämnesområde (t.ex. "nanoteknik").
- Olika referensformat t.ex. BibTeX, EndNote.

7.4. Administration av persondata

Många universitet använder redan i nuvarande lokala databaser ett autentiseringssystem, vilket kopplar till lokala persondata. I nuläget finns dock inget system för att på nationell nivå ta fram ett unikt ID.

För vissa typer av publikationer kan det kanske vara möjligt att utgå från det system som redan finns i Libris, och sedan koppla till informationen från de lokala databaserna i SwePub.

8. Utvärderingsgränssnitt

8.1. Översikt och innehåll

Då utvärdering bygger på att aggregerade data kan sammanställas och visas på ett lättöverskådligt sätt kommer det att krävas en del utvecklingsarbete av ett analysverktyg för att åstadkomma detta.

Utgångspunkten bör vara att all data klassificerad på miniminivå enligt SVEP-MODS i de lokala publiceringsdatabaserna ingår i underlaget för analyserna. Ingångssidan skall ligga utanför Libris webbsök under www.swepub.se och det ska på ingångssidan klart och tydligt ges en detaljerad information om vad innehållet i de lokala databaserna består av, vilka lärosäten som ingår, tidsperioder för respektive lärosäte, vilka publikationstyper som ingår samt definitionen för dessa, kriterier för bedömning av refereegranskning etc. Även eventuella påpekanden om brister i underlaget bör påpekas här.

Åtkomst till analysverktyget skall vara fri liksom möjligheterna för export.

8.2. Relation till övriga system

Utgångspunkten är att grunddata ligger i Libris databas och att analysverktyget arbetar mot denna och att inte någon parallell databas byggs upp. Analysverktyget måste dock kunna sammanställa informationen från den gemensamma basen och kunna uttrycka detta i numeriskt format.

Fråga att lösa: Hur ska beräkningsdata föras över till analysverktyget? Denna fråga skall lösas under projektiden.

Eftersom det i analysverktyget endast görs enklare statistiska beräkningar krävs att man kan exportera data till andra program eller applikationer. Ett lämpligt format för export är tabbseparerat textformat. Även XML, JDCB samt OAI-PMH bör erbjudas (se 10.2.2). Visningsformatet skall vara HTML.

8.3. Principiella funktioner

Analysverktyget skall vara uppbyggt så att ett antal fördefinierade rapporter genereras utifrån ett val av parametrar gjorda av frågeställaren. Dessa rapporter kan t ex ge uppgifter om antal publikationer per lärosäte i tidsserier, med möjlighet till vidare uppdelning på publikationstyper, enhet, språk, ämne etc.

Vid beräkning på högsta aggregerade nivå ska hänsyn tas till dubbelräkning, d v s om en publikation har författare från två olika institutioner ska endast en publikation anges i lärosätets (eller vetenskapsrådets, fakultetens) totalsiffra.

8.3.1. Ingående parametrar i analysverktyget

- Enhet (endast lärosäte, institution)
- Publikationstyp
- Språk
- Innehåll (refereegranskning, populärvetenskap etc.)
- Tidsperiod
- Ämne (enligt SVEP:s indelning)¹⁰

¹⁰ I aktörsintervjuerna är frågan om ämnesklassificering den mest svårtolkade frågan. Där uttrycks allt från starkt positiv inställning (VR:s analysenhet) till direkt avvisande (RUT 2, flera forskarröster). Förutom frågan om hur SVEP:s ämnesindelning ser ut och hur acceptansen för denna är måste man vid analyser ta hänsyn till lokala strukturer, t ex grad av specialisering på ett lärosäte.

Ämne förslås att vara med bland parametrarna men bör användas med stor försiktighet vilket också ska framgå på webbplatsen.

8.3.2. Förslag till ytterligare parametrar utifrån aktörsintervjuerna

För att analyser utifrån dessa parametrar ska vara möjliga att göra krävs att data bearbetas i de lokala databaserna eller i andra program/applikationer. Implementeringen av dessa parametrar görs därför efter att detta skett.

Tidskrifter

(Kräver en omfattande identifiering av titlar och ISSN-nummer. Analyser utifrån tidskriftsklasser kan vara ett alternativ till ämnesutvärdering).

Fraktionerade författardata

(Kräver en bearbetning av data av något slags statistikprogram, t.ex. Excel, ProDiver etc.)

9. Icke-grafiska exportgränssnitt

Det finns önskemål om export av hela eller uttag av delar av SwePub. Följande är hämtade från inventeringsfasen.

9.1. OAI-PMH

För export till söktjänster som OAISTER¹¹, DART¹², forskning.se¹³ etc. Möjlighet att kunna leverera sets som enbart poster med fri fulltext och efter dokumenttyp, ämne, lärosäte, enbart kvalitetskontrollerade.

Exportformat minst dc:simple¹⁴ (ev. med DRIVER¹⁵ anpassning) samt SwePub MODS.

9.2. Excel (csv)

Möjlighet att ta ut en sökning i administrativt gränssnitt som tabbseparerad fil för konvertering till Excel

¹¹ Se: <http://www.oiaster.org/>

¹² Se: <http://www.dart-europe.eu/>

¹³ Se: <http://forskning.se/>

¹⁴ Se: <http://dublincore.org/documents/dces/>

¹⁵ Se: <http://www.driver-support.eu/en/guidelines.html>

9.3. Andra format

Andra format som SUHF-Metricsgruppen önskar sig för utsökning är JDBC¹⁶, XML, OAI-PMH (SwePubs MODS-profil) samt hela databasen som en textfil.

9.4 Google (och andra sök-tjänster)

Optimal Google-anpassning. Indexering i Google Scholar.

¹⁶ Se: http://en.wikipedia.org/wiki/Java_Database_Connectivity

10. Projektbeskrivning och tidplan

Projektarbetet kring utvecklingen av databasen är uppdelat i sju arbetspaket (AP), varib arbetet med kravspecifikation utgör AP2.

AP	Namn	Projektid
1	Projektledning	sept 07-aug 09
2	Utveckling av kravspecifikation	sept 07-juni 08
3	Databasimplementering	sept 08-aug 09
4	Förstudie till nationell lösning för universitet utan publiceringsdatabas	apr 08-dec 08
5	Information och marknadsföring	sept 08-aug 09
6	Utveckling av analysverktyg för databasen	okt 08-aug 09
7	Test och utvärdering	sept 08-aug 09

11. Styrorganisation

Projektgruppen kommer att avsluta sitt arbete i oktober 2009 och den nuvarande styrgruppens mandatperiod avslutas också under hösten 2009. I vårt uppdrag ingår att förbereda projektets övergång till ordinarie verksamhet och att ge förslag till styrorganisation för denna.

11.1. Under projektet

Under projektets gång leds projektet av en styrgrupp bestående av

- Gunilla Herdenberg, chef för Avdelningen för nationell samverkan, KB (ordf.)
- Jan Hagerlid, Avdelningen för nationell samverkan, KB (sekr.)
- Lars Björnshauge, universitetsbibliotekarie vid Lunds universitet
- Ulf Göranson, överbibliotekarie vid Uppsala universitet
- Agneta Olsson, överbibliotekarie vid Göteborgs universitet
- Eva Müller, chef för Avdelningen för digitala Biblioteket, KB
- (+ efterträdare för Lisa Petersen)

11.2. I driftsläget

SUHF förväntas få ett stark inflytande när den nya styrorganisationen ska träda i kraft. I rapporten Kartläggning av högskolans elektroniska publicering och rekommendationer för nationell samverkan föreslås att SUHF tar ansvar för en permanent övergripande nationell samverkan kring elektronisk publicering och

publiceringsredovisning.¹⁷ I Danmark finns förhoppningar att SUHF:s danska motsvarighet Danske Universiteter¹⁸ kommer att ansvara för den danska forskningsdatabas som nu utvecklas vid Forsknings- og Innovationsstyrelsen.

Styrorganisationen kan formeras som en samverkansgrupp med representanter för SUHF, KB, universitets- och högskolebiblioteken, Vetenskapsrådet, Vetenskapsakademien och KK-stiftelsen.

När det gäller ansvar för drift- och utveckling fastställdes i projektutlysningen att "söktjänsten ska utvecklas och tas i permanent drift av avdelningen för Libris".

11.3 Från projekt till kontinuerlig drift

Det är mycket viktigt att det blir en omedelbar övergång från projekt till ordinarie verksamhet i oktober 2009. Styrgruppen och projektgruppen har då fullgjort sina uppdrag. Systemägarfrågan och finansieringen skall vara klar och det ska finnas ett förslag till en samverkansgrupp.

¹⁷ Kartläggning av högskolans elektroniska publicering och rekommendationer för nationell samverkan, Arbetsgruppen för högskolans elektroniska publicering på uppdrag av SUHF:s Forum för bibliotekschefer. Januari 2008. Rapporten samt presentation för SUHF:s Forum för bibliotekschefer är tillgängliga från <http://www.suhf.se/>

¹⁸ "Danske Universiteter" är de danska universitetens intresseorganisation som behandlar "uddannelses -og forskningspolitiske spørgsmål", <http://www.rektorkollegiet.dk>

Bilaga SwePub kravspecifikation

Formatspecifikation

Stefan Andersson,

Digital publicering,
Uppsala universitetsbibliotek

Draft
Mars 2008

Inledning

Denna formatspecifikation anger hur metadataposter i lärosätenas lokala publikationsdatabaser skall göras tillgängliga för de nationella tjänster som utvecklas i projektet *Samlad ingång till och redovisning av svensk vetenskaplig publicering*¹⁹.

Specifikationen utgår från de rekommendationer till *Nationellt format för publiceringsdatabaser (lokala register över akademisk publicering)*²⁰ som godkändes av SUHF och publicerades i augusti 2005.

Först listas de metadataelement som är obligatoriska. I de obligatoriska elementen kan även finnas underelement som inte behöver tas med men som rekommenderas. Dessa listas för varje element under rubriken *Tillägg*. De icke obligatoriska element som sedan listas behöver inte ingå i metadataposten men rekommenderas om de är tillämpliga.

Några tillägg och förtydliganden av formatet har gjorts i förhållande till de ursprungliga rekommendationerna:

Översikt över ändringar

Upphovsman

- Roller anges som standard med koder från *marcrelator*²¹.

¹⁹ <http://www.kb.se/om/projekt/openaccess/projekt/#Samlad>

²⁰ http://epc.ub.uu.se/files/publ1_2.pdf

²¹ <http://www.loc.gov/marc/sourcecode/relator/relatorsource.html>

- Valfritt antal hierarkiska nivåer kan anges för ansvarig institution men högsta nivån (lärosäte) samt institution eller motsvarande är fortfarande obligatoriska.

Publikationstyp

- Nya publikationstyper:
 - artikel, forskningsöversikt (tidigare recension)
 - proceedings (redaktörsskap)
 - samlingsverk (redaktörsskap)
- Innehållsmärkning görs till ett eget element som kan kombineras enligt tabellen på sid. 6-7.

Värdepublikation

- ISSN för tidskrifter som värdepublikation görs obligatoriskt.

Identifikatorer för elektroniska dokument

- Alla metadataposter skall ha en länk till motsvarande post i de lokala databaserna.

Beskrivning av annan medietyp

- Beskrivningsstandarden skall vara enligt MODS²².

ISBN

- Görs obligatoriskt (om det finns).

Språk

- Görs obligatoriskt. Standarden skall vara iso639-2b²³.

²²

<http://www.loc.gov/standards/mods/v3/mods-userguide-elements.html#typeofresource>

²³ <http://www.loc.gov/marc/languages/>

Metadadataformat

Allmänt

- Metadatatposter skall hämtas från de lokala publiceringsdatabaserna via OAI-PMH²⁴.
- Det skall vara möjligt att hämta endast de poster som är publicerade genom det egna lärosätet, dvs. att upphovsmannen var knuten till lärosätet vid publiceringstillfället
- Metadadataformatet skall vara MODS²⁵, version 3.2 eller högre.
- Alla element (fält) kan upprepas valfritt antal gånger.
- Element utan innehåll skall inte medtagas i beskrivningen.
- Teckenkodningen skall vara UTF-8.
- Om språkattributet *lang* används skall språket anges enligt iso639-2b²⁶.

Obligatoriska metadataelement

Upphovsman

Med upphovsman menas person eller institution (hit räknas även exempelvis kongress eller forskningsprojekt) som är ansvarig för det intellektuella eller konstnärliga innehållet i ett dokument.

²⁴ <http://www.openarchives.org/OAI/openarchivesprotocol.html>

²⁵ <http://www.loc.gov/standards/mods/>

²⁶ <http://www.loc.gov/marc/languages/>

Upphovsmannen kan vara antingen författare eller redaktör, rollen skall specificeras i metadata. Rollen skall anges med koder från *marcrelator*²⁷. Obligatoriska är *aut* (författare) resp. *edt* (redaktör/utgivare).

För personnamn anges både för- och efternamn. En rekommendation är att dessa anges i separerade fält. Om detta inte är möjligt måste de anges enligt modellen Efternamn, Förnamn i ett *namePart*-element utan attribut.

Författare anges i den ordning de står i publikationen.

Adressangivelser

Adress (dvs. lärosäte plus institution eller motsvarande) måste anges för åtminstone lärosätets egna författare (och helst för samtliga författare).

Adressen skall anges enligt modellen *Lärosäte, Institution el. motsv.*

Tillägg

- Relevanta årtal (födelsedatum) kan läggas till personlig upphovsman.
- Upphovsmansinformationen kan knytas till ett lokalt auktoritetsregister. I *authority*-attributet anges då lärosätets domän eller annan unik kod och i *xlink:href*-attributet anges en URI eller annat unikt ID till den lokala posten.

Exempel

```
<name type="personal" authority="uu" xlink:href="N2-1690">
  <namePart type="given">Leif</namePart>
  <namePart type="family">Svensson </namePart>
  <namePart type="date">1963-</namePart>
  <role>
    <roleTerm type="code" authority="marcrelator">aut</roleTerm>
  </role>
  <affiliation>Uppsala universitet, Litteraturvetenskapliga
institutionen</affiliation>
</name>
```

²⁷ <http://www.loc.gov/marc/sourcecode/relator/relatorsource.html>

Ansvarig organisation

Den organisatoriska enhet som upphovsmannen tillhör vid det egna lärosätet, och som är knuten till publiceringen, skall anges i ett separat element. De olika hierarkiska nivåerna skall läggas i olika underelement från den högsta till den lägsta. Lärosätet är alltid den högsta nivån och måste alltid anges. Nivåer motsvarande fakultet eller institution skall också anges. Rollen för ansvarig organisation skall anges med koden *pbl* från *marcrelator*. Lärosätets domän eller annan unik kod anges i *authority*-attributet.

Tillägg

- Om organisationsnamnen även finns i engelska versioner kan ett separat element läggas till för dessa. *Lang*-attributet är då obligatoriskt.
- Organisationsinformationen kan knytas till ett lokalt auktoritetsregister via *xlink:href*-attributet.

Exempel

```
<name type="corporate" lang="swe" authority="uu" xlink:href="uu-XX">
  <namePart>Uppsala universitet</namePart>
  <namePart>Humanistisk-samhällsvetenskapliga
vetenskapsområdet</namePart>
  <namePart>Historisk-filosofiska fakulteten</namePart>
  <namePart>Litteraturvetenskapliga institutionen</namePart>
  <role>
    <roleTerm type="code" authority="marcrelator">pbl</roleTerm>
  </role>
</name>
```

Titel

Ord eller fras som namnger ett verk. Huvudtitel och undertitel läggs i separata underelement. Om detta inte är möjligt skall titeln anges enligt modellen Huvudtitel : undertitel i *title*-elementet.

Alternativa titlar anges i separat element med värdet *alternative* i *type*-attributet.

```
<titleInfo>
  <title>Twenty years to nowhere</title>
  <subTitle>does land use affect regeneration perspectives?</subTitle>
```

</titleInfo>

<titleInfo lang="eng" type="alternative ">

<title>Variations in Russian Verbal Government 1800–1840</title>

</titleInfo>

Publikationstyp

Publikationen skall klassificeras enligt följande typologi:

<i>Svenska</i>	<i>Engelska</i>	<i>Kod</i>	<i>Ref.</i>	<i>Övr. vet.</i>	<i>Övr. pop.</i>
artikel	article	art	X	X	X
artikel, forskningsöversikt	article, review/survey	for	-	X	-
artikel, recension	article, book review	res	-	X	X
doktorsavhandling	doctoral thesis	dok	-	X	-
licentiatavhandling	licentiate thesis	lic	-	X	-
bok	book	bok	-	X	X
kapitel	book chapter	kap	-	X	X
konferensbidrag	conference paper	kon	X	X	X
proceedings (redaktörsskap)	conference proceedings	pro	-	X	X

konstnärlig forsknings- och utvecklingsrapport	artistic research and development report	kfu	-	X	X
rapport	report	rap	-	X	X
samlingsverk (redaktörsskap)	collection (editor)	sam	-	X	X
övrigt	other	ovr	-	X	X

Ovanstående tabell anger vilka möjliga värden som finns för innehållsmärkning (X = möjlig).

Innehållsmärkning

<i>Svenska</i>	<i>Engelska</i>	<i>Kod</i>
refereegranskat	refereed	ref
övrigt vetenskapligt	other scientific	vet
övrigt (populärvetenskap, debatt)	other (popular scientific, debate etc.)	pop

Både publikationstyp och innehållsmärkning måste alltid anges enligt koder i ovanstående tabell och följande modell:

Exempel

```
<genre authority="svep" type="publicationType">art</genre>
<genre authority="svep" type="contentType">ref</genre>
```

Definitioner av publikationstyper

Artikel

Avser en artikel i vetenskaplig eller annan tidskrift.

Recensioner avser recension av bok (book review) i tidskrift eller dagstidning

Forskningsöversikt avser review/survey i vetenskapliga tidskrifter.

Bok

Monografisk publikation bestående av en eller flera fysiska delar, avsett att utgöra en avslutad helhet. Omfattar även e-bok.

Kapitel i bok eller rapport

Självständig del av monografisk publikation. Ingår ofta i samlingsverk, ett antal fristående verk som utgivits tillsammans eventuellt under redaktionell styrning till ett visst tema.

Doktorsavhandling

Avser godkänd avhandling för doktorsexamen.

Licentiatavhandling

Avser godkänd avhandling för licentiatexamen.

Konferensbidrag

Referee granskat konferensbidrag avser sådant bidrag som på förhand granskats av en eller flera referenter och som utges i officiell proceeding eller konferensmeddelande.

Övriga konferensbidrag avser bidrag som publiceras i konferensskrift men som ej genomgått förhandsgranskning.

Proceedings (redaktörsskap)

Avser eget redaktörsskap för officiell konferensproceeding eller konferensmeddelande. Egna enskilda bidrag redovisas under *Konferensbidrag*.

Rapport

Avser publikation ingående i rapportserie.

Samlingsverk (redaktörsskap)

Avser publikation där ett antal var för sig fristående verk utgivits tillsammans. Endast redaktören skall redovisas under denna kategori. De enskilda verken skall var för sig redovisas som *Kapitel* eller del av bok.

Konstnärlig forsknings- och utvecklingsrapport

Rapport från ett konstnärligt forsknings- och/eller utvecklingsprojekt. En sådan rapport kan ta sig uttryck även i andra medietyper än text. Detta anges då under rubriken *Beskrivning av annan mediatyp*.

Övrigt

Avser publikationer där ingen annan publikationstyp är tillämplig.

Definitioner av innehållsmärkning

Referegranskat

Avser att publikationen genomgått granskning av oberoende referenter vars samlade bedömning avgör om publikationen uppfyller tillräckliga krav på vetenskaplig kvalitet för att accepteras för publicering. Gäller oftast artiklar men även konferensbidrag kan vara referegranskade.

Övrigt vetenskapligt

Avser att publikationens innehåll är av vetenskaplig karaktär och riktar sig mot vetenskapssamhället men ej genomgått referegranskning.

Övrigt (populärvetenskap, debatt)

Avser att publikationens innehåll är av populärvetenskaplig karaktär och riktar sig mot allmänheten. Även debattinlägg räknas hit.

Utgivningsuppgift

Utgivningsår anges alltid. Med utgivningsår menas det år ett arbete publiceras och att man härvidlag följer förlagens utgivningstid. T.ex. skall en artikels upphovsår anges till det år tidskriften publiceras och inte det år artikeln accepteras.

Exempel

```
<originInfo>  
  <dateIssued>1994</dateIssued>  
</originInfo>
```

Tillägg

För monografier bör även utgivningsort och förlag anges.

Exempel

```
<originInfo>  
  <place>  
    <placeTerm>Uppsala</placeTerm>  
  </place>  
  <publisher>Institutionen för cell- och molekylärbiologi</publisher>  
  <dateIssued>1994</dateIssued>  
</originInfo>
```

Ämnesområde

Angivande av ämnesområde enligt rekommendationer för *Ämneskategorier för vetenskapliga publikationer*²⁸ från DP 1 inom SVEP. Dessa bygger på Statistiska centralbyråns, SCB:s *Nationell förteckning över forskningsämnen*²⁹ samt Vetenskapsrådet utvidgade klassificering³⁰. SVEP-kategorierna skall anges enligt nedanstående modell (svensk och engelsk form).

Exempel

```
<subject lang="swe" authority="svep" xlink:href="112104">  
  <topic>HUMANIORA och RELIGIONSVETENSKAP</topic>  
  <topic>Estetiska ämnen</topic>  
  <topic>Litteraturvetenskap</topic>
```

²⁸ Se: <http://www.ub.uu.se/epub/categories/>

²⁹ Se: http://www.scb.se/templates/Standard_24458.asp

³⁰ Se: <http://www.snac.vr.se/classifications.htm>

```
</subject>
<subject lang="eng" authority="svep" xlink:href="112104">
  <topic>HUMANITIES and RELIGION</topic>
  <topic>Aesthetic subjects</topic>
  <topic>Literature</topic>
</subject>
```

Om det inte är möjligt att medtaga hela hierarkin kan kategorierna anges på en nivå enligt modellen:

Exempel

```
<subject lang="swe" authority="svep" xlink:href="112104">
  <topic>Litteraturvetenskap</topic>
</subject>
<subject lang="eng" authority="svep" xlink:href="112104">
  <topic>Literature</topic>
</subject>
```

Tillägg

Fria ämnesord kan anges med eller utan *lang*-attribut.

Exempel

```
<subject lang="eng">
  <topic>Hermeneutics</topic>
</subject>
```

Värdpublikation

Med värdpublikation menas en seriell eller monografisk publikation varur endast en mindre, självständig del utgör underlag för en referens, t.ex. en tidskriftsartikel eller ett kapitel i en bok. Värdpublikationens titel och siduppgifter anges samt, i förekommande fall vid artiklar, volym, häfte och tidskriftens ISSN.

Rekommendation: Utskriven titel föredras framför förkortningar vid angivande av tidskriftstitel.

Exempel

```
<relatedItem type="host">
  <titleInfo>
    <title>Journal of Applied Ecology</title>
  </titleInfo>
  <identifier type="issn">1057-9478</identifier>
```

```
<part>
  <detail type="volume">
    <number>31</number>
  </detail>
  <detail type="issue">
    <number>3</number>
  </detail>
  <extent>
    <start>11</start>
    <end>23</end>
  </extent>
</part>
</relatedItem>
```

Identifikatorer för elektroniska dokument

Alla metadatatposter skall ha en länk till motsvarande post i de lokala databaserna. Om lärosätet själv publicerar dokumentet elektroniskt skall adressen (URL) för publiceringen anges.

Rekommendation: Om möjligt anges även ett s.k. DOI-nummer (Digital Object Identifier)³¹ vilken gör att publikationen kan spåras även om den byter publiceringsadress.

Även post-ID:n från andra externa databaser kan anges: PubMed, ISI eller Scopus. Enbart numret anges, ej inledande prefix som doi: eller ISI:.

Tillåtna värden i *type*-attributet är uri, doi, hdl, pmid, isi eller scopus.

Exempel

```
<identifier type="uri">http://opus.uu.se/publication.xml?id=103993</identifier>
```

En URL till den lokala posten. Obligatorisk.

```
<identifier type="doi">10.1000/182</identifier>
```

Ett DOI-nummer.

```
<identifier type="hdl">1721.1/5759</identifier>
```

Ett handle-nummer.

```
<identifier type="pmid">121668435</identifier>
```

Ett PubMed-nummer.

³¹ Se: <http://www.doi.org/>

```
<identifier type="isi">000249875900007</identifier>
```

Ett ISI-nummer.

```
<identifier type="scopus">12345678900</identifier>
```

Ett Scopus-nummer.

URL:er till fulltexter eller andra tillhörande filer anges i elementet *location/url*. En förklarande text kan läggas i *displayLabel*-attributet.

```
<location>  
  <url displayLabel="Fulltext, fritt tillgänglig">http://...</url>  
</location>
```

Om fulltexten dessutom är publicerad i ett lokalt öppet arkiv skall följande element (för utsökning av sådana) tillfogas posten:

```
<genre authority="svep" type="linkType">fulltext</genre>
```

ISBN

ISBN (internationellt standardnummer för böcker). Anges alltid om det finns.

Exempel

```
<identifier type="isbn">91-554-6028-3</identifier>
```

Språk

Uppgift om dokumentets språk. Anges alltid enligt enligt iso639-2b.

Exempel

```
<language>  
  <languageTerm type="code" authority="iso639-2b">eng</languageTerm>  
</language>
```

Ej obligatoriska metadataelement

Beskrivning av annan medietyp

Om andra medietyper än text tas med i det lokala systemet kan dessa anges enlighet med MODS³².

Exempel

```
<typeOfResource>still image</typeOfResource>
```

Abstract

Ett referat av dokumentets innehåll. Språkattribut kan läggas till vid behov.

Exempel

```
<abstract lang="eng"> ...Text...</abstract>
```

Anmärkningar

Eventuella anmärkningar rörande dokumentet. Språkattribut kan läggas till vid behov.

Exempel

```
<note lang="eng">...Text...</note>
```

Fysisk beskrivning

Antal sidor eller annan fysisk beskrivning av dokumentet (gäller monografiska publikationer).

Exempel

```
<physicalDescription>  
  <extent>30 s.</extent>  
</physicalDescription>
```

³² <http://www.loc.gov/standards/mods/v3/mods-userguide-elements.html#typeofresource>

Serie

Uppgift om dokumentet ingår i en serie. Serietitel, ISSN och del anges alltid i förekommande fall.

Exempel

```
<relatedItem type="series">
  <titleInfo>
    <title>Comprehensive Summaries of Uppsala Dissertations from the
      Faculty of Science and Technology</title>
  </titleInfo>
  <identifier type="issn">1104-232X</identifier>
  <identifier type="issue number">1008</identifier>
</relatedItem>
```

Relaterade publikationer

Uppgift om relaterade dokument, exempelvis olika versioner/format eller recensioner. Anges enligt MODS i *relatedItem*-elementet. Alla ovanstående element kan ingå. Följande varianter finns:

Exempel

```
<relatedItem type="otherVersion">
  <!--Innehåll, annan upplaga med reviderat/ändrat innehåll-->
</relatedItem>

<relatedItem type="otherFormat">
  <!--Innehåll, samma innehåll, annat format (tryckt, elektroniskt)-->
</relatedItem>

<relatedItem type="isReferencedBy">
  <!--Innehåll, refereras av -->
</relatedItem>

<relatedItem type="constituent">
  <!--Innehåll, del av-->
</relatedItem>
```


Utkast för remiss

Detta är ett utkast till kravspecifikation för SwePub-databasen. Remissvar ska vara projektgruppen tillhanda på e-post eller vanlig post senast 2008-05-15.

E-post: Jorgen.Eriksson@lub.lu.se

Vanlig post:

Projektgruppen för SwePub

c/o Jörgen Eriksson

Biblioteksdirektionen, Lunds universitet

Box 134

221 00 LUND