

SUHF Arbetsgrupp för vidareutveckling av format- specifikationen för SwePub MODS på nationell nivå

Slutrapport

SUHF:s arbetsgrupp Maj 2012

Jörgen Eriksson,
Lunds universitet

Stefan Andersson,
Uppsala universitet

Peter Hansson,
Chalmers tekniska högskola

Ulf Kronman,
Kungliga Tekniska högskolan

Jessica Lindholm,
Malmö högskola

Johan Toll,
Linköpings universitet

Per Åberg,
Kungl. biblioteket

Innehåll

1.Uppdraget	2
1.2 Projekt mål.....	3
1.3 Aktiviteter	3
2. Sammanfattning av förslaget.....	4
2.1 Arbetsgruppen föreslår följande förändringar i formatet.	4
3. Förslag till fortsatt skötsel av SwePubs MODS-format	6
4. SwePub formatbeskrivning. Version 1.2 2012	7
4.1 Översikt över tillägg och ändringar i version 1.2, 2011	9
5. Riktlinjer för tillämpningar av SwePub formatspecification 1.2, 2012	10
6. Refereegranskat eller inte?.....	24
7. Bilaga: Projektplan	26

1. Uppdraget

SUHF:s förbundsårsamling fattade vid Förbundsårsamlingen i Eskilstuna den 28 oktober beslut om att anta rekommendationerna REK2010:4 med anledning av rapporten "Kvalitet och publikationsdatabaser". Ett av delbesluten innebär att SUHF initierade en vidareutveckling av formatspecifikationen för SwePub MODS på nationell nivå genom att en nationell samordningsgrupp utsågs. Kungl. Biblioteket (KB) har en arbetsgrupp för infrastruktur och användartjänster inom ramen för programmet OpenAccess.se. För att möjliggöra en samordning mellan SUHF:s grupp och KB:s grupp innehöll den beslutade arbetsgruppen två personer ur KB:s arbetsgrupp.

Gruppens ordförande föreslog en projektplan (Bilaga 1) som stämde av med ledningen för Forum för Bibliotekschefer och godkändes av SUHF:s generalsekreterare. Planen tog sikte på en slutrapportering i mars 2012.

SUHF:s presidium utsåg följande arbetsgrupp

Jörgen Eriksson, Lunds universitet, sammankallande
Johan Toll, Linköpings universitet
Stefan Andersson, Uppsala universitet
Peter Hansson, Chalmers
Jessica Lindholm, Malmö högskola
Per Åberg, Kungl. Biblioteket (KB utsåg Kristin Olofsson som senare ersattes av Per Åberg)
Ulf Kronman, Kungliga tekniska högskolan (tillkom på gruppens initiativ)

1.2 Projekt mål

Projektet skall leverera

- ett uppdaterat SwePub utbytesformat (av de fält som är av intresse för analyser) med sikte på att skapa förbättrade analysmöjligheter med jämförbar data.
- en praktisk handledning som beskriver vad som krävs för att höja datakvaliteten i de lokala databaserna
- förslag på fortsatt skötsel av formatet

1.3 Aktiviteter

Arbetsgruppen har haft tre arbetsmöten. I Lund 17 maj, i Stockholm 5 september 2011 och i Stockholm 26 januari 2012.

I oktober 2011 skickades det framtagna förslaget ut för kommentarer till SUHF:s Forum för bibliotekschefer och till ansvariga för publikationsdatabaser vid lärosätena.

Efter att svaren samlats in och sammanställts utgjorde de grunden för en workshop "Kvalitetssäkring av publikationsdatabaser" som hölls på Kungliga Biblioteket i Stockholm 29/11. Workshopen blev givande, med 40 deltagare, och gav ytterligare input till arbetsgruppen utöver de tidigare kommentarerna. Rapporten skickades ut för en sista kommentarrunda första veckan i maj till ansvariga för publikationsdatabaser.

2. Sammanfattning av förslaget

SwePub skall ge en samlad ingång till forskningspublikationer från svenska lärosäten. Poster som beskriver publikationer hämtas från lärosätenas lokala publikationsdatabaser varje natt och görs sökbara i ett gemensamt gränssnitt. För att en post skall tas med i SwePub så måste den validera, dvs vissa fält måste finnas med i posten. Tjänstens främsta syfte är att öka synlighet och tillgänglighet till svenska forskningspublikationer och valideringen är anpassad därefter.

Idag levererar 35 lärosäten poster till SwePub som 2012-04-20 innehåller ca 475600 bibliografiska poster. SwePub driftas av LIBRIS-avdelningen vid Kungliga Biblioteket.

Vi föreslår att SwePub som informationskanal skall finnas kvar med oförändrat syfte och fortsätter att valideras på samma sätt som nu. Till det läggs ytterligare ett valideringssteg. Poster som godkänns i den andra valideringen hamnar i en delmängd av SwePub som kan användas för analyser enligt vårt förslag. Med en sådan modell kan SwePub som informationsspridare fortsätta att växa och de poster som levererats tidigare finnas kvar i tjänsten, även om de inte är godkända för analys. Existerande poster kan sedan uppgraderas till att klara analysnivån efter hand vid varje lärosäte.

Förslaget innebär att de av SwePub insamlade bibliografiska posterna på sikt omvandlas till en så kallad "databrunn", där SwePub:s nuvarande söktjänst och en framtida möjlig analystjänst utgör två olika användartjänster som utnyttjar olika urval dataposter från samma källa, beroende på hur posterna är validerade och uppmärkta.

2.1 Arbetsgruppen föreslår följande förändringar i formatet.

- Alla poster i SwePub ska kunna kopplas till minst ett svenskt lärosäte eller annan offentlig forskningsinstitution på ett kontrollerat sätt via auktoriserade organisationsidentifikatorer.
- Alla poster ska kopplas till minst den högsta nivån i ämnesområde enligt Standard för svensk indelning av forskningsämnen 2011[8].
- Författare som tillhör det lärosäte som lagt in posten ska kunna identifieras med organisationstillhörighet via auktoriserade organisationsidentifikatorer.
- Uppgift om det totala antalet författare till en publikation ska finnas i samtliga poster.

- Posterna ska innehålla ett fält som anger om lärosätet godkänt att posten används för analys. Fältet är tänkt att kunna användas när/om SwePub-data börjar användas för bibliometriska analyser och eventuell medelsfördelning.
- Möjlighet att märka fler publikationstyper med innehållstypen "refereegranskat" för att öka möjligheterna till förfinade analyser.

Om dessa förändringar genomförs blir det möjligt att genomföra bibliometriska studier med författarfraktionering av artiklarna per lärosäte och sedan göra författarfraktionerade jämförelser mellan lärosätena, vilket inte är möjligt idag.

Samtliga lärosäten som kommenterade förslaget (15 st) svarade att de skulle kunna leverera information enligt ovan redan nu eller senast under 2013.

Frågor av betydelse för att få kvalitetssäkrad, analyserbar data som inte löses av ändringar i formatspecifikationen är de som är av bedömningstyp, framför allt vad som skall räknas till en viss publikationstyp och vad som skall räknas som refereegranskat. Vi öppnar för en diskussion om detta i appendix "Refereegranskat eller inte?". Ett annat problemområde som ligger utanför formatet är forskare som är knutna till flera lärosäten. När skall en publikation kopplas till det ena eller det andra lärosätet? Här är det viktigt att lärosätena kan nå en samsyn om det skall gå att ta fram jämförbara analyser. Därför föreslår vi följande för att föra det här arbetet vidare, som för övrigt kommer att behöva revideras och förändras löpande beroende på nya krav på analys och förändringar i publiceringssätten. Förslaget till förvaltning utvecklas i nästa stycke efter punktlistan.

Arbetsgruppen föreslår följande vad gäller fortsatt skötsel och utveckling

- Att det tillsätts en arbetsgrupp under SUHF:s Forum för bibliotekschefer med uppgift att koordinera samverkan mellan de lokala publikationsdatabaserna och vidareutveckling av riktlinjer.
- Att den löpande driften och förvaltningen av SwePub sköts av Kungliga Bibliotekets LIBRIS-avdelning som idag och att även förvaltningen av MODS-formatet läggs på samma huvudman.
- Att förändringarna i formatet och organisationen implementeras med sikte på att vara genomförda under 2013.

3. Förslag till fortsatt skötsel av SwePubs MODS-format

I arbetsgruppens uppdrag har ingått att ta fram ett förslag till fortsatt skötsel av det utbytesformat – SwePub MODS – som används vid överföring av data från svenska organisationers¹ publikationsdatabaser till Kungliga Bibliotekets SwePub-system.

Arbetsgruppen bedömer att MODS XML-format som sådant inte behöver någon specifik organisation för löpande underhåll. De förändringar av formatet som föranleds av en framtida utveckling av användningen av universitetens publikationsdatabaser och SwePub kan istället hanteras med ad hoc-insatser initierade av den grupp för samverkan kring riktlinjer för datainmatning som föreslås nedan. Arbetsgruppen har som en del av projektet uppdaterat och tagit fram nya riktlinjer för inmatning av publikationsdata i universitetens publikationsdatabaser, med syfte att erhålla så enhetliga och analyserbara data som möjligt i det gemensamma systemet SwePub. Riktlinjerna är dock inget statistiskt dokument som löser t.ex. avvägningsproblem rörande olika publikationstyper och författaraffilieringar en gång för alla, utan de kommer att behöva förankras, diskuteras och revideras löpande.

Speciellt påtagligt kommer behovet av en revision att vara om data från SwePub kommer att användas i ett system för medelsfördelning till lärosätena i enlighet med förslag i den aktuella utredningen *Prestationsbaserad resurstilldelning för universitet och högskolor*.²

Arbetsgruppen föreslår därför att det skapas en grupp för löpande samverkan kring riktlinjerna för inmatning av data i universitetens publikationsdatabaser. Då inmatningen av data i publikationsdatabaserna i stort sett uteslutande administreras av personal vid landets högskole- och universitetsbibliotek ser projektgruppen det som lämpligt att denna samordningsgrupp skapas som ytterligare en arbetsgrupp underställd SUHF:s Forum för bibliotekschefer.

Då inmatningen av data i publikationsdatabaserna är en angelägenhet för ett stort antal bibliotek vid landets lärosäten behövs sannolikt en konstruktion med en central arbetande grupp med representativ bemanning, kompletterad med ett större nätverk av bibliotekspersonal som arbetar med datainmatning i sitt dagliga arbete. Konstruktionen kan liknas vid och modelleras efter den samverkan som tidigare byggts upp kring katalogisering och inmatning av data i bibliotekens samkatalog LIBRIS.

Arbetsgruppen ser vidare att den löpande driften och förvaltningen av SwePub och MODS-formatet lämpligen sköts av Kungliga bibliotekets LIBRIS-avdelning. LIBRIS-avdelningen bör också kunna sköta viss teknisk koordinering av de register som kommer att behövas för att erhålla enhetliga och analyserbara data i det gemensamma publikationsdatabassystemet.

Delar av de riktlinjer som arbetats fram hanteras och utvecklas av andra, varför samverkan och löpande informationsinhämtning med externa intressenter bör rymmas inom den föreslagna arbetsgruppens

¹ Organisationer som levererar publikationsdata till SwePub förtecknas här <http://www.swepub.se/help.jsp>

² Från regeringen.se: Rapporten *Prestationsbaserad resurstilldelning för universitet och högskolor*. Diarienummer U2011/7356/UH. Remiss utdatum 2011-12-23. Sista svarsdatum 2012-03-15.

fokus. Det kan gälla aktiviteter på Högskoleverket kring de nationella ämneskategorierna³ och att löpande säkra att den svenska användningen av MODS inte avviker från de user guidelines som Library of Congress utarbetar och underhåller⁴.

4. SwePub formatbeskrivning. Version 1.2 2012

Inledning

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

Under våren 2008 skickades kravspecifikationen för SwePub, inklusive formatspecifikationen, på remiss till samtliga medlemmar i SUHF samt ytterligare några aktörer. Utifrån de remissvaren gjordes sedan ett antal ändringar och tillägg.

Under år 2011-12 har formatet åter reviderats av *SUHF:s arbetsgrupp för vidareutveckling av formatspecifikationen för SwePub MODS på nationell nivå*.

Rotelement

Element: <mods>

Ansvarigt lärosäte

Element: <recordInfo>

Underelement: <recordContentSource>, attribut: authority

Upphovsman

Element: <name>, attribut: type, authority

Underelement: <namePart>, attribut: type

Underelement: <role> med underelement:<roleTerm>, attribut: type, authority

Underelement: <affiliation>

Underelement: <description>

Ansvarig organisation

Element: <name>, attribut: type, lang, authority

Underelement: <namePart>

³ Standard för svensk indelning av forskningsämnen 2011.

<http://www.hsv.se/download/18.cb396a51350f671dad80001160/standard120126.pdf>

⁴ <http://www.loc.gov/standards/mods/mods-guidance.html>

Underelement: <role> med underelement:<roleTerm>, attribut: type, authority
Underelement: <description>

Titel

Element: <titleInfo>, attribut: lang, type
Underelement: <title>
Underelement:<subTitle>

Publikationstyp och innehållsmärkning

Element: <genre>, attribut: authority, type

Utgivningsuppgift

Element: <originInfo>
Underelement: <dateIssued>
Underelement: <place> med underelement: <placeTerm>
Underelement: <publisher>

Ämnesområde

Element: <subject>, attribut: lang, authority, id
Underelement: <topic>

Värdpublikation

Element: <relatedItem>, attribut: type
Underelement: <titleInfo> med underelement: <title>
Underelement: <identifier>, attribut: type
Underelement: <part> med underelement <detail>, attribut: type med underelement <number> samt underelement <extent> med underelement: <start> och underelement: <end>

Identifikatorer

Element: <identifier>, attribut: type
Element: <location>
Underelement: <url>, attribut: note, displayLabel
Underelement: <physicalLocation>, attribut type, authority

Språk

Element: <language>
Underelement: <languageTerm>, attribut: type, authority

Annan medietyp

Element: <typeOfResource>

Abstract

Element: <abstract>, attribut: lang

Anmärkningar

Element: <note>, attribut: lang

Antal upphovsmän

Element: <note>, attribut: type

Godkänd för analys

Element: <note>, attribut: type

Fysisk beskrivning

Element: <physicalDescription>

Underelement: <extent>

Serie

Element: <relatedItem>, attribut: type

Underelement: <titleInfo> med underelement: <title>

Underelement: <identifier>, attribut: type

Relaterade publikationer

Element: <relatedItem>, attribut: type

Underelement: alla ovanstående

4.1 Översikt över tillägg och ändringar i version 1.2, 2011

Version av MODS

- Ändrat till version 3.4 eller högre.

Upphovsman

- Olika personliga ID:n från lokala och externa databaser eller system kan anges på ett enhetligt sätt.
- Adress till lärosätets egna författare måste anges.
- Lärosätets internetdomän skall anges inom klammer efter namnet i adressen.

Publikationstyp

- Det är nu möjligt att lägga innehållsmärkning refereegranskad på fler publikationstyper.

Ämnesområde

- Nationella ämneskategorier skall anges enligt HSV:s nya förteckning (hsv2011) istället för de gamla SVEP-kategorierna.

Identifikatorer för elektroniska dokument

- Det är nu möjligt att beskriva vilken url (om flera har specificerats) som pekar mot det lokala arkivet.

Anmärkningar

- Antal författare totalt för en publikation skall anges.
- Möjligt att ange att en post har godkänts/kvalitetssäkrats lokalt, för att kunna användas vid analys (används ej ännu).

5. Riktlinjer för tillämpningar av SwePub formatspecification 1.2, 2012

Detta är en beskrivning av formatets praktiska tillämpning, med XML-exempel.

Först listas allmänna rekommendationer kring hur formatet ska användas. Därefter beskrivs varje huvudelement i formatet och hur det ska tillämpas. Obligatoriska fält har markerats med en asterisk (*) i rubriken. I de obligatoriska elementen kan det även finnas underelement som inte behöver tas med men som rekommenderas. Dessa listas för varje element under rubriken Tillägg. De icke obligatoriska element som listas behöver inte ingå i metadataposten men rekommenderas om de är tillämpliga.

Allmänt

- Metadataposter skall hämtas från de lokala publiceringsdatabaserna via OAI-PMH[2].
- 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.
- Poster som beskriver (ännu) ej publicerade publikationer skall inte exporteras till SwePub.
- Metadataformatet skall vara MODS[3], version 3.4 eller högre.
- Alla huvudelement (fält) förutom recordInfo och physicalDescription kan upprepas valfritt antal gånger.
- Element utan innehåll skall inte tas med i beskrivningen.
- Teckenkodningen skall vara UTF-8.
- Om språkattributet lang används skall språket anges enligt iso639-2b[4].
- Om lang-attributet inte används skall fältvärdet vara på svenska.
- Om ett fält anges på flera språk skall lang-attributet alltid användas.
- Element som innehåller en sträng t. ex ISSN eller DOI skall inte innehålla någon ytterligare information eller inledande och avslutande mellanslag, dubbla mellanslag eller kontrolltecken.
- Exempel: <identifier type="issn">1057-9478</identifier> är korrekt men inte <identifier type="issn">1057-9478 (e-issn) </identifier>

Metadataelement

Rotelement (*)

Rotelement som också anger vilka XML-scheman som används ska anges i början av varje post enligt följande.

```
<mods xmlns="http://www.loc.gov/mods/v3"
version="3.4"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xlink="http://www.w3.org/1999/xlink"
xsi:schemaLocation="http://www.loc.gov/mods/v3 http://www.loc.gov/standards/mods/v3/mods-3-4.xsd">
```

Ansvarigt lärosäte (*)

Ansvarigt lärosäte (där posten hämtas ifrån) anges som metametadata enligt följande. Lärosätets id baseras på internetdomänen. I de fall flera lärosäten levererar data till SwePub tillsammans så anges varje lärosäte under varsitt element recordContentSource. Listan över möjliga värden fastställs av SwePubs förvaltningsorganisation.

Exempel: Uppsala universitet som äger internetdomänen uu.se

```
<recordInfo>
  <recordContentSource authority="swepub.se">uu.se</recordContentSource>
</recordInfo>
```

Exempel: Chalmers och Göteborgs universitet med internetdomänerna chalmers.se och gu.se som levererar data tillsammans

```
<recordInfo>
  <recordContentSource authority="swepub.se">chalmers.se</recordContentSource>
  <recordContentSource authority="swepub.se">gu.se</recordContentSource>
</recordInfo>
```

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. Upphovsmannen kan vara antingen författare eller redaktör, rollen skall specificeras i metadata. Rollen skall anges med koder från *marcrelator*[6]. 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 *skall* anges i den ordning de står i publikationen.

Adressangivelser (*)

Adress (dvs. lärosäte plus institution eller motsvarande) måste anges för lärosätets egna författare (och helst för samtliga författare). Tillsammans med anmärkningen *Antal upphovsmän* blir det möjligt att identifiera vilka författare och hur många som hör till ett lärosäte av det totala antalet, vilket är nödvändigt för att kunna göra bibliometriska analyser som använder sig av författarfraktionering.

Adressen skall anges enligt modellen *Lärosäte [internetdomän]; <Orgenhet [internetdomän:id];> Institution el. motsv. [internetdomän:id]*. Högsta nivån ska vara lärosätet och lägsta nivån ska vara institution eller motsvarande. Om tillämpligt rekommenderas att fler nivåer mellan högsta och lägsta nivån anges, t ex fakultet. Varje nivå separeras med semikolon. Efter det egna lärosätets namn skall motsvarande domän anges inom klammer (obligatoriskt). Efter övriga egna enheter anges, om möjligt, lärosätets domän följt av kolon och dess lokala ID inom klammer. Uppgiften skall gälla den adress som är kopplad till själva publikationen, ej den adress som upphovsmannen för närvarande är kopplad till – om den inte är densamma som i publikationen.

Eventuella externa adresser skall inte märkas upp med domän eller ID inom klammer. För svenska externa lärosäten rekommenderas den namnform som används av SUHF[5]

Tillägg

- Födelseår kan läggas till personlig upphovsman.
- Upphovsmansinformationen kan knytas till ett auktoritetsregister. I authority-attributet anges då lärosätets domän eller annan unik kod som benämner auktoritetsregistret.
- Personliga identifikatorer i lokala och externa databaser eller system kan läggas till i description-elementet enligt modellen "internetdomän=ID".

Exempel på författare från det egna lärosätet som är kopplad till ett auktoritetsregister

```
<name type="personal" authority="uu.se">
  <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 [uu.se]; Litteraturvetenskapliga institutionen
[uu.se:123456]</affiliation>
  <description>uu.se=N2-1690</description>
  <description>orcid.org=1234123412341234</description>
  <description>researcherid.com=123456</description>
  <description>scopus.com=123456</description>
</name>
```

Exempel på extern författare som saknar koppling till ett auktoritetsregister

```
<name type="personal">
  <namePart type="given">Jannike</namePart>
  <namePart type="family">Johansson</namePart>
  <namePart type="date">1972</namePart>
  <role>
    <roleTerm type="code" authority="marcrelator">aut</roleTerm>
  </role>
  <affiliation>Stanford University; Department of Chemistry</affiliation>
  <description>orcid.org=1234123412341234</description>
</name>
```

Exempel på affiliering med institution tillhörande två lärosäten från samma publikationsdatabas

```
<affiliation>Chalmers tekniska högskola [chalmers.se]; Institutionen för matematiska vetenskaper [chalmers.se:371]</affiliation>
```

```
<affiliation>Göteborgs universitet [gu.se]; Naturvetenskapliga fakulteten [gu.se:65]; Institutionen för matematiska vetenskaper [gu.se:371]</affiliation>
```

Exempel på flerspråkig affiliering

```
<affiliation lang="swe">Göteborgs universitet [gu.se]; IT-fakulteten [gu.se:121]; Institutionen för tillämpad informationsteknologi [gu.se:1293]</affiliation>
```

```
<affiliation lang="eng">University of Gothenburg [gu.se]; IT Faculty [gu.se:121]; Department of Applied Information Technology [gu.se:1293]</affiliation>
```

Antal upphovsmän (*)

Det totala antalet upphovsmän för en publikation **måste** anges i särskild anmärkning med attributet *nameCount*.

Exempel

```
<note type="nameCount">5</note>
```

Ansvarig organisation (planeras att avvecklas) (*)

Om möjligt ska i första hand adressangivelser till upphovsman tillhörande det egna lärosätet ske under fältet *Upphovsman*. På sikt när samtliga lärosäten kan koppla adress till de lokala författarna i fältet *Upphovsman* kommer fältet *Ansvarig organisation* att ersättas av dessa adressangivelser.

Den organisatoriska enhet som upphovsmannen tillhör vid det egna lärosätet, och som är knuten till publiceringen, skall anges i ett separat obligatoriskt 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*.

Om en publikation har upphovsmän från flera olika institutioner eller enheter vid det egna lärosätet skall alla dessa enheter anges i separata element med lärosätet som högsta nivå.

Organisatoriska förändringar hanteras i de lokala systemen.

Ansvarig organisation skall inte blandas ihop med utgivare/förlag. I de fall ansvarig organisation och utgivare/förlag är samma enhet kan uppgiften anges i båda elementen.

Tillägg

- Om organisationsnamnen även finns i engelska versioner kan ett separat element läggas till för dessa. Lang-attributet är då obligatoriskt.
- Identifierare för angiven organisatorisk enhet, som används i lokala och externa databaser eller system, kan läggas till i description-elementet enligt modellen "internetdomän=ID".

Exempel

```
<name type="corporate" lang="swe" authority="uu.se">
  <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>
  <description>uu.se=123</description>
</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.

Exempel på titel utan språk angivet

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

Exempel på flera titlar på olika språk

```
<titleInfo>
  <title lang="swe">Artikeltitel</title>
  <subTitle lang="swe">Artikels undertitel</subTitle>
</titleInfo>
<titleInfo>
```

<title lang="eng">Article Title</title>

<subTitle lang="eng">Article Subtitle</subTitle>

</titleInfo>

Exempel på en alternativtitel

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

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

</titleInfo>

Publikationstyp (*)

Publikationen skall klassificeras enligt följande typologi:

Typ	Kod	Ref. vet.	Övr. vet.	Övr. pop.
artikel	art	X	X	X
artikel, forskningsöversikt	for	X	X	-
artikel, recension	rec	X	X	X
doktorsavhandling	dok	-	X	-
licentiatavhandling	lic	-	X	-
bok	bok	X	X	X
kapitel i bok eller rapport	kap	X	X	X
konferensbidrag	kon	X	X	X
patent	pat	-	-	-
proceedings (redaktörskap)	pro	-	X	X
rapport	rap	X	X	X
samlingsverk (redaktörskap)	sam	-	X	X
konstnärligt arbete	kfa	X	X	X
övrigt	ovr	X	X	X

Ovanstående tabell anger vilka möjliga värden som finns för innehållsmärkning (X = möjlig). Innehållstyp måste kombineras med publikationstyp och endast en av vardera ovanstående skall ges till varje publikation.

Innehållsmärkning (*)

Typ	Innehåll	Kod
-----	----------	-----

refereegranskat	Riktat sig mot vetenskapssamhället	ref
övrigt vetenskapligt	Riktat sig mot vetenskapssamhället	vet
övrigt (populärvetenskap, debatt)	Riktat sig mot allmänheten	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, antologi 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

Refereegranskat 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 (redaktörskap)

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

Rapport

Avser publikation t.ex. ingående i rapportserie.

Samlingsverk (redaktörskap)

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.

Konstnärligt arbete

Konstnärlig forskning tar sin utgångspunkt i den konstnärliga processen och verksamheten. Forskningen, som kan beröra alla konstarter, är praktikbaserad och inkluderar en intellektuell reflektion för att utveckla ny kunskap. Resultatet från konstnärlig forskning redovisas vanligen i både gestaltande och skriftlig form[7].

Övrigt

Avser publikationer där ingen annan av ovanstående publikationstyper är tillämplig, t.ex. poster (presentationer), evenemang, konserter, utställningar, teaterföreställningar eller textilier.

Definitioner av innehållsmärkning

Refereegranskat

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 och andra publikationstyper 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.

Se appendix för diskussion kring innehållsmärkning.

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 eller ändras till det år tidskriften publiceras och inte det år artikeln accepteras eller görs tillgänglig som "ahead-of-print" eller "corrected proof".

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 (*)

Ämnesområde enligt Standard för svensk indelning av forskningsämnen 2011[8] som är framtagen av SCB och Högskoleverket skall anges på valfri nivå.

Exempel

```
<subject lang="eng" authority="hsv2011" id="60302">
  <topic>Humanities</topic>
  <topic>Philosophy, Ethics and Religion</topic>
  <topic>Ethics</topic>
</subject>
<subject lang="swe" authority="hsv2011" id="60302">
  <topic>Humaniora</topic>
  <topic>Filosofi, etik och religion</topic>
  <topic>Etik</topic>
</subject>
```

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

Exempel

```
<subject lang="eng" authority="hsv2011" id="60302">
  <topic>Ethics</topic>
</subject>
<subject lang="swe" authority="hsv2011" id="60302">
  <topic>Etik</topic>
</subject>
```

Tillägg

- Fria ämnesord eller nyckelord kan anges med eller utan lang-attribut. Detta element kan användas för att lägga till ämnesord som saknas bland de kategorierna.

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. Flera ISSN (print/electronic) kan anges om så önskas.

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>
```

Volym, häfte och sidor ska anges i separata fält, så långt det är möjligt. Om inte detta går praktiskt, så kan följande format användas.

Exempel

```
<part>
  <detail type="citation">
    <caption>Vol. 117, nr. 6, s. 30-38</caption>
  </detail>
</part>
```

Identifikatorer för elektroniska dokument (*)

Alla metadataposter 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)[9] 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 bestäms av SwePubs förvaltning och är vid version 1.2 fastställande uri, doi, hdl, pmid, isi eller scopus.

Exempel

En URL till den lokala posten. Obligatorisk.

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

Ett DOI-nummer.

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

Ett handle-nummer.

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

Ett PubMed-nummer.

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

Ett ISI-nummer.

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

Ett Scopus-nummer.

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

URL:er till fulltexter eller andra tillhörande filer kan anges i elementet *location/uri*. En valfri förklarande text kan läggas i *displayLabel*-attributet. Om fulltexten är fritt tillgänglig används *note="free"*.

```
<location>
```

```
  <uri note="free" displayLabel="Publikation, fritt tillgänglig">http://...</uri>
```

```
</location>
```

Ytterligare exempel

```
<location>
```

```
  <uri note="free" displayLabel="Publikation, förlagsversion">http://...</uri>
```

</location>

Om fulltexten är publicerad i ett lokalt öppet arkiv bör följande konstruktion användas:

```
<location>
  <url note="free" displayLabel="Fulltext, fritt tillgänglig">http://...</url>
  <physicalLocation type="electronicRepository" authority="swepub.se">uu.se</physicalLocation>
</location>
```

ISBN (*)

ISBN (internationellt standardnummer för böcker). Anges alltid om det finns tilldelat publikationen. För att underlätta maskinell matchning ska endast själva ISBN-numret anges.

Exempel

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

Språk (*)

Uppgift om dokumentets språk. Anges alltid enligt enligt iso639-2b. För icke-språkligt material kan språkkoden "zxx" användas. Om språk är okänt använd språkkod "und".

Exempel

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

Beskrivning av annan medietyp

Om andra medietyper än text tas med i det lokala systemet kan dessa anges enligt med MODS[10].

Exempel

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

Abstract

Ett referat av dokumentets innehåll. Språkattribut kan läggas till vid behov. HTML-kod är ej tillåtet i abstract-fältet.

Exempel

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

Anmärkningar

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

Exempel

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

Godkänd för analys

Markerar att en post har godkänts/granskats lokalt för användning vid analys. Elementets innehåll består av en siffra som motsvarar en viss bestämd godkännandenivå. OBS! Nationella riktlinjer för hur detta fält ska användas saknas.

Exempel

<note type="approvedForAnalysis">1</note>

Fysisk beskrivning

Antal sidor eller annan fysisk beskrivning av dokumentet.

Exempel

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

Exempel

```
<physicalDescription>
  <extent>1 bordsduk : blå- och vitrutig ; 90 x 90 cm</extent>
</physicalDescription>
```

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>

<relatedItem type="references">
<!-- enligt MODS 3.4 spec: Information concerning a resource cited or referred to in the resource. -->
</relatedItem>

<relatedItem type="reviewOf">
<!-- enligt MODS 3.4 spec: Information concerning a resource reviewed in the content of the resource. -
->
</relatedItem>

[1] <http://swepub.kb.se>

[2] <http://www.openarchives.org/OAI/openarchivesprotocol.html>

[3] <http://www.loc.gov/standards/mods/>

[4] <http://www.loc.gov/marc/languages/>

[5] <http://www.suhf.se/web/Members.aspx> .

[6] <http://www.loc.gov/marc/relators/>

[7] <http://www.vr.se/forskningvistodjer/forskningsomraden/konstnarligforskningochutveckling.4.3376a56c12918b8d17b80001186.html>

[8] <http://app.ub.uu.se/epub/categories/>

[9] <http://www.doi.org/>

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

6. Refereegranskat eller inte?

Vad en publikation klassificeras som har också betydelse för hur den räknas i bibliometriska analyser. För att posterna i SwePub skall kunna fungera som ett rättvisande underlag för jämförelser mellan lärosäten behövs en samsyn vad gäller klassificering av dokument som refereegranskade, övrigt vetenskapligt eller populärvetenskapligt. Den här texten är att se som en start på en diskussion med syfte att nå en samsyn hos våra publikationsdatabaser.

Möjligheten att klassa flera publikationstyper (bok, bokkapitel,...) som refereegranskade än tidigare ingår i det här förslaget med samma definition som för de övriga publikationstyperna.

Refereegranskning "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."

Vi föreslår att de oberoende referenterna också skall vara externa, alltså att de inte tillhör samma lärosäte som författaren.

Krav på refereegranskning av böcker och bokkapitel innebär i praktiken att så gott som samtliga böcker som publiceras på nationella förlag skall placeras i kategorin "Övrigt vetenskapligt" eftersom refereegranskning i princip inte tillämpas av svenska förlag.

För tidskrifter kan Ulrichs International Periodicals Directory fungera som ett riktmärke. Om en tidskrift anges "Refereed/peer reviewed" där så kan artiklar från tidskriften klassas som Refereegranskad. Liksom om en tidskrift indexerats av Thomson Reuters, Scopus eller PubMed.

Konferenspublikationer upplever vi som särskilt svåra att gradera. Här följer ett förslag som kommer från det OA.se-finansierade projektet "Parallellpublicering av konferensartiklar (2012)".

"När man tittar i olika publikationsdatabaser kan man se att papers från samma konferenspublikation kan ha olika klassningar vad gäller publikationstyp och/eller kvalitetsgranskning. Därför vill vi betona vikten av att vi registrerar publikationer på samma vis. Här följer vårt förslag till dylika riktlinjer.

- Opublicerade konferensbidrag: Publikationsdatabaserna skall kunna användas även för publicering av konferensbidrag som inte publiceras i de vanliga kanalerna (proceedings, tidskrifter, antologi eller book of abstracts). Detta kan göras med hjälp av en samlingsklassificering som konferenspresentationer i det lokala systemet. Hit hör också posters, PPT-presentationer o. dyl. Alternativt att man har (en del av) de olika publikationsformerna som egna publikationstyper i sitt institutionella arkiv, med möjlighet att märka upp om en publikation är publicerad eller inte. Hamnar under "Övrigt" i SwePub.
- Abstracts. Book of abstracts är en sammanställning av sammanfattningar av presentationer vid en konferens. Dessa är ofta kvalitetsgranskade. Även om ett abstract är kvalitetsgranskat är det långt ifrån säkert att den presentation som den hänger samman med är det. I så fall skall

abstractet registreras som kvalitetsgranskat, men inte fullängdspappret. Abstract kan klassificeras under samlingstypen konferenspresentationer i det lokala systemet om man så vill. Abstracts skall inte sammanblandas med fullängdsartiklar från en konferens. Hamnar under "Övrigt" i SwePub.

- Proceedings: sammanställning av fullängdspapper från en konferens. Enskilda artiklar skall ha publikationstyp konferensbidrag. Många gånger är det svårt att avgöra om ett konferensbidrag har omarbetats och genomgått kvalitetsgranskning för senare publicering som tidskriftsartikel eller bokkapitel. Avgörande för om ett konferensbidrag skall klassas om är att den har genomgått en omarbetning och ev. kvalitetsgranskning. Exempel på en serie som ibland kan innehålla omarbetade artiklar, ibland inte, är "Lecture Notes in Computer Science". Serien innehåller ofta utökade konferensbidrag och vad som särskiljer den artikel som klassas som tidskriftsartikel eller konferensbidrag är inte alltid klart. Ett alternativ är att alltid gå på klassningen i ISI i det här fallet.
- Bokkapitel. Inom framför allt samhällsvetenskap och humaniora kan konferenser resultera i att bidragen publiceras, efter viss redigering och mer eller mindre noggrann kvalitetsgranskning, i tryckta antologier. Detta uttrycks ofta med en fras liknande "This **edited** volume is the outcome of two workshops convened at the ..."
"Collection of **Revised** Papers originally read at the International Conference ..."
Sådana publikationer bör klassas som kapitel i bok/antologi.
- Tidskriftsartiklar. Om det inte uttryckligen står i tidskriften att detta är ett nummer dedicerat till konferensartiklar eller om det står att konferensartiklarna har omarbetats för tidskriftspublicering skall de anses som kvalitetsgranskade tidskriftsartiklar om det framgår av tidskriftens hemsida att granskningsprocedur förekommer. Om en publikation är klassad både som konferensartikel och tidskriftsartikel skall den klassas som tidskriftsartikel."

Vi hoppas på en fortsatt diskussion som leder till ökad samsyn kring frågan refereegranskat eller inte och vad som hör hemma i de olika publikationstyperna.

7. Bilaga: Projektplan

Projekt mål

Projektet skall leverera

ett uppdaterat SwePub utbytesformat (av de fält som är av intresse för analyser) med sikte på att skapa förbättrade analysmöjligheter med jämförbar data.

en praktisk handledning som beskriver vad som krävs för att höja datakvalitén i de lokala databaserna
förslag på fortsatt skötsel av formatet

Uppdaterat SwePub utbytesformat med handledning

- Arbetsgruppen utgår från de rekommendationer som finns i avsnitt 4 och 5 i rapporten "Kvalitet och publikationsdatabaser" och tar fram ett konkret förslag på ändringar i SwePub-formatet.
- Förslaget skickas på remiss till ansvariga för publiceringsregistreringstjänster vid samtliga lärosäten. För varje Swepub-fält och dess föreslagna användning ber vi dem som svarar att tala om vilka möjligheter de har att uppfylla kraven och om de inte klarar det nu inom vilken tidsram de ev. kan tänkas göra det. Samt fråga hur fältet skapas och hur innehållet kontrolleras idag.
- Avsluta remissrundan med en workshop där alla intresserade kan delta.
- Föreslå SUHF en uppdaterad version av SwePub utbytesformat och en tidsplan för införandet (ev. förändringar i söktjänsten SwePub borde synkas med en sådan tidsplan).
- Föreslå uppdaterade anvisningar till SwePub utbytesformat och göra det fastställda formatet tillgängligt via swepub.kb.se
- Ta fram ett förslag till det långsiktiga vårdandet av SwePubs utbytesformat. Det borde finnas någon organisation som tar ansvar för att formatet sköts om kontinuerligt i framtiden.
- Avrapportering