

Behov och rekommendation kring format i Swepub

Sammanfattning

Detta dokument är framtaget för att kunna användas som beslutsunderlag gällande överföringsformatet för dataleveranser till Swepub. I dagsläget används Swepub MODS som bygger på MODS-formatet utvecklat av Library of Congress. MODS är i grunden ämnat för bibliografiska data men har inom ramen för projektet Vidareutveckling av Swepub utvecklats för att lämpa sig även för bibliometriska data. Detta har lett till vissa avvikelser från det ursprungliga MODS-formatet som försvårar återanvändning av data. Dokumentet beskriver de utvecklingsbehov gällande metadatakrav som har identifierats för att kunna använda Swepub som källa både bibliografiska och bibliometriska underlag och hur MODS alternativt andra format kan besvara dem. Utöver MODS har BIBFRAME, CERIF och Dublin Core undersökts.

BIBFRAME har potential att vara framtida format för Swepub eftersom det stödjer länkade data och används som grund i Libris XL men eftersom det fortfarande är under utveckling är det än så länge för tidigt att börja anpassa BIBFRAME som Swepubs överföringsformat för bibliometriska behov. CERIF är utvecklat som format för forskningsinformation men har inte den bibliografiska beskrivningen som utgångspunkt som Swepub har. Dublin Core är etablerat och användbart men skulle behöva modifieras för de bibliometriska behoven så som MODS. Library of Congress utvecklar fortfarande MODS aktivt och den senaste utvecklingen har tillfört ändringar i formatet som motsvarar behoven i Swepub. Formatbyte är ett omfattande arbete som behöver tillföra väsentlig nytta för att vara värt besväret för KB och Swepub-organisationerna. Det behöver även tas hänsyn till synnergieffekter med Libris-infrastrukturen vilket i det långa loppet kan påverka det interna formatet i Swepub-systemet. Rekommendationen är därmed att inte byta ut överföringsformatet för tillfället utan att fortsätta använda MODS tills vidare och att vidareutveckla Swepub MODS utifrån de senaste versionerna av MODS.

Bakgrund

Metadatastandarder för informationsutbyte inom forskningens infrastruktur behövs för att forskningsinformation ska kunna delas och återanvändas, både manuellt och maskinellt. För väl fungerande metadatastandarder behöver intressenterna inom forskningsinfrastrukturen enas om vilken typ av information som efterfrågas, hur den ska vara strukturerad och hur överföringen ska ske. För detta behöver bland annat enas om överföringsformatet.¹

För dataleveranser till Swepub används i dagsläget överföringsformatet MODS². MODS är schema för bibliografiska dataelement utvecklat och upprätthållen av Network Development and MARC Standards Office vid Library of Congress. Inom *Samordning av den svenska högskolans elektroniska publicering* (SVEP), som hade som uppdrag att utarbeta ett förslag till gemensamma riktlinjer för

¹ Riktlinjer för utbyte av forskningsinformation 1.0.2. 2015, s. 6. <https://www.swerim.se/riktlinjer-och-standarder/informationsutbyte/>

² MODS Metadata Object Description Schema: <http://www.loc.gov/standards/mods/>

publiceringsdatabaser vid svenska lärosäten, valdes MODS som format 2005.³ Projektet för *Samlad ingång till och redovisning av svensk vetenskaplig publicering* utgick från dessa riktlinjer när Swepub MODS 1.0 framställdes som överföringsformat för den nationella söktjänsten för forskningspublikationer Swepub 2008.⁴

Sveriges universitets- och högskoleförbund (SUHF) initierade 2011 en arbetsgrupp för vidareutveckling av formatspecifikationen för Swepub MODS för att göra data analysbara för bibliometriska jämförelser på nationell nivå och öka datakvaliteten på lärosätenas publikationsdatabaser. I sin slutrapport föreslog arbetsgruppen hur Swepub MODS formatet behövde utvecklas för analyserbara data.⁵ Inom projektet *Vidareutveckling av Swepub* har formatet Swepub MODS utvecklats utifrån förankrade behov hos Vetenskapsrådet (VR) och SUHF.⁶ Följande bild beskriver den underliggande datamodellen för forskningsinformation som Swepub samlar in, skapar relationer emellan och lämnar ut i dataunderlag utifrån de förankrade behoven:

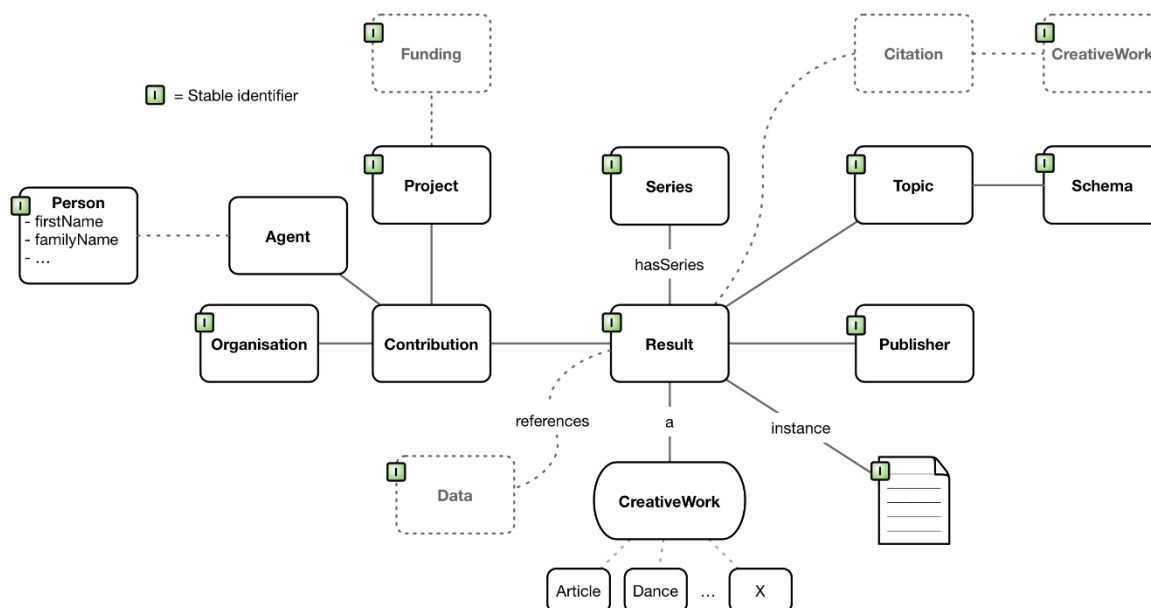


Bild: Martin Malmsten, KB, 2018

Sedan 2017 har Kungliga biblioteket fått anslagshöjning för att bland annat vidareutveckla Swepub ytterligare.⁷ Under 2017 har arbete pågått i Gruppen för Swepub under *Forum för nationell bibliotekssamverkan och utveckling* för att ta fram strategisk riktning för vidareutveckling av Swepub.⁸ Innan projektet *Vidareutveckling av Swepub* avslutades januari 2017 för att gå över till den löpande verksamheten sammanställdes *Rekommendationer och förslag på åtgärder för vidareutveckling av*

³ Rekommendationer : Nationellt format för publiceringsdatabaser. 2005, s. 2.

http://www.kb.se/Dokument/Om/projekt/SVEP/nationellt_format.pdf

⁴ Swepub – formatspecifikation 1.0. 2008, s. 2.

[http://www.kb.se/dokument/SwePub/\[v.1.0\]%20Swepub%20MODS%20formatspec.pdf](http://www.kb.se/dokument/SwePub/[v.1.0]%20Swepub%20MODS%20formatspec.pdf)

⁵ SUHF arbetsgrupp för vidareutveckling av formatspecifikationen för Swepub MODS på nationell nivå. 2012.

<http://bit.ly/1cPRDb>

⁶ SwePub MODS metadata format specification 2.6. 2015. [http://www.kb.se/dokument/SwePub/v.-2.6-SwePub MODS_Final_version_2015_09_10.pdf](http://www.kb.se/dokument/SwePub/v.-2.6-SwePub_MODS_Final_version_2015_09_10.pdf)

[http://www.kb.se/dokument/SwePub/v.-2.6-SwePub MODS_Final_version_2015_09_10.pdf](http://www.kb.se/dokument/SwePub/v.-2.6-SwePub_MODS_Final_version_2015_09_10.pdf)

⁷ Regleringsbrev för budgetåret 2017 avseende Kungl. biblioteket.

<https://www.esv.se/statsliggaren/regleringsbrev/?RBID=17964>

⁸ Forum för nationell bibliotekssamverkan och utveckling, Gruppen för Swepub.

<http://www.kb.se/bibliotek/Forum-for-nationell-bibliotekssamverkan-och-utveckling/SwePub/>

*Swepub*⁹ där en av punkterna var att se över formatet. Denna formatutredning är framtagen för att kartlägga behoven i Swepub och tillämpningsmöjligheterna i det befintliga formatet samt överväga alternativ till MODS-format. Utredningen är gjord av Tuija Drake med input hämtad från Miriam Björkhem, Stina Degerstedt, Ulf Kronman, Camilla Lindelöw, Martin Malmsten och Bodil Wennerlund från KB, Lia Fallenius, Lena Hagberg, Marie Lövgren, Joakim Philipson, Mia Sandström och Anders Söderbäck från Stockholms universitetsbibliotek, Urban Andersson, Jessica Lindholm och Kristin Olofsson från Chalmers tekniska högskolas bibliotek samt Jörgen Eriksson, Mikael Graffner och Hanna Voog från Lunds universitetsbibliotek.

Syfte

Överföringsformatet MODS som används för närvarande för dataleveranser till Swepub är i grunden ämnat för bibliografiska data. Inom ramen för projektet *Vidareutveckling av Swepub* har formatet vidareutvecklats så att det lämpar sig även för bibliometriska data. Detta har krävt vidareutveckling av formatet som avviker från det ursprungliga MODS-formatet och försvårar återanvändning av data. Ytterligare stöd för bibliometriska analyser, uppföljning, och återanvändning av metadata kan kräva flera avvikelser. Därför är det viktigt att se över Swepub MODS-format om det går att utveckla utan att avvika för mycket från det ursprungliga MODS-formatet, alternativt överväga ett byte av överföringsformat.

VR har tagit fram *beställarunderlag för ett analyserbart Swepub*¹⁰ vilket kommer att användas som underlag för vidareutvecklingen av Swepub. I det fortsatta arbetet kommer VR:s kravspecifikation användas som utgångspunkt till datakvalitetshöjning i Swepub. Under det arbetet behöver det säkerställas att formatet som används för dataleveranser till Swepub motsvarar den funktionaliteten som behövs för analyserbart Swepub.

Behov

Om datakvaliteten i Swepub uppnår en tillräckligt god kvalitetsnivå avser VR att använda Swepub för två syften i första hand: för det första bibliometriska analyser och för det andra import av publikationer till ansökningshanteringssystemet Prisma. Nedan beskrivs dessa behov uppdelade i respektive syfte och kvarstående behov av åtgärder för att kunna uppfylla kravspecifikationen.

Metadatakrav för bibliometriska analysunderlag

Följande uppgifter är avgörande i medräkning i bibliometriska analyser och behöver vara kvalitetssäkrade:

- ISSN/ISBN
- Publiceringsstatus
- Utgivningsdatum
- Digital Object Identifier (DOI)/Swepub-ID
- Affiliering
- Totalt antal upphovspersoner
- Totalt antal lokala upphovspersoner uppdelat på kön
- (Ämnesklassning på nivå 2 (3-siffernivå) enligt standarden av SCB/UKÄ¹¹)

⁹ Rekommendationer och förslag på åtgärder för vidareutveckling av SwePub. 2017.

https://www.kb.se/dokument/SwePub/SwePub_rekommendationer_20170130_FINAL.pdf

¹⁰ Beställarunderlag för ett analyserbart Swepub. 2017.

<http://www.kb.se/dokument/SwePub/Gruppen%20f%C3%B6r%20SwePub/Best%C3%A4llarunderlag%20f%C3%B6r%20ett%20analyserbart%20SwePub.pdf>

¹¹ Standard för svensk indelning av forskningsämnen 2011. <http://www.scb.se/dokumentation/klassifikationer-och-standarder/standard-for-svensk-indelning-av-forskningsamnen/>

En auktoriserad lista över publiceringskanaler är en förutsättning för att kunna avgöra vad för forskningsoutput som medräknas i bibliometriska analyser. För att kunna koppla forskningsoutput och dess publiceringskanaler mot auktoriserade listor behöver data från de lokala publikationsdatabaserna innehålla kvalitetssäkrade ISSN- och ISBN-nummer samt Digital Object Identifier (DOI). I de fall där det saknas en sådan identifikator alternativt då identifikatorn angiven i outputen är ogiltig, skulle man kunna överväga att tillåta andra typer av identifikatorer, till exempel LibrisID, samt införa möjligheten att registrera identifikatorer som ogiltiga. Ett alternativ är även att införa SwepubID. Lärosätena förespråkar detta eftersom det skulle skilja posterna i de lokala publikationsdatabaserna från de berikade posterna i Swepub, i stället för att använda samma ID som i de ursprungliga posterna.

Publiceringsstatus och utgivningsdatum behövs för avgränsning i dataunderlaget. Det behöver dock tas ställning till output som tilldelas DOI och publiceras som preprint och online-version först före den officiellt publicerade versionen och som börjar redan då samla på sig citeringar. *Nationella riktlinjer för beskrivning av vetenskaplig output* rekommenderar ändring av publiceringsstatus och utgivningsdatum i en befintlig post eftersom detta påverkar dubbelthanteringen i systemet¹² men det går eventuellt att införa ytterligare fält för andra utgivningsdatum kopplade till olika skeden av publicering.

Antalet alla och lokala upphovspersoner samt deras affilieringar används i fraktionering av andelar av en output. VR är även intresserad av att få med uppgiften om kön för de lokala upphovspersonerna. Det behöver dock diskuteras vidare om bibliografiska data över forskningsoutput ska innehålla könsrelaterade data över upphovspersoner, det vill säga hur känsligt är sådana data över personer i en öppen, nationell databas. Affilieringar är vidareutvecklade i Swepub MODS 2.6 eftersom MODS-formatet inte stödjer den typen av affilieringar som behövs för att validera kopplingen mellan lokala upphovspersoner och registrerande organisation. För att koppla ihop namn och tillhörande organisation skall den lokala personIDn därmed anges tillsammans med den registrerande organisationens URI bestående av domännamn. Utifrån de angivna lokala namnen räknas antalet lokala upphovspersoner medan antalet alla upphovspersoner registreras som en siffra. Trots att det kan vara svårt att koppla samman personID och affiliering till den registrerande organisationen i ett renodlat MODS-format så bör dock angivning av ORCID¹³ som personID ses över utifrån det nyintroducerade subelementet *nameIdentifier* i MODS 3.6¹⁴.

Antalet upphovspersoner räknas av systemet utifrån angivna roller enligt standarden *MARC Code List for Relators*¹⁵. Roller författare ("aut"), redaktör ("edt") har definierats som upphovspersoner per automatik och andra roller, som översättare, illustratör eller regissör, kan förses med en ytterligare roll creator ("cre") för att markeras som upphovsperson och därmed räknas med i analysunderlagen. Det saknas dock vissa typer av konstnärliga roller och roller som är vanligt förekommande inom forskningsinformation som corresponding author, det vill säga författaren som är ansvarig för dialogen mellan förlaget och författarna.

Även ämnesklassning på nivå 2 (3-siffernivå) utifrån *Standard för svensk indelning av forskningsämnen 2011* utarbetad av Statistiska Centralbyrån (SCB) och Högskoleverket (HSV), numera Universitetskanslersämbetet (UKÄ), avses vara viktigt för bibliometriska analyser som utgår från ämne. I Swepub saknar ca tredjedel av alla poster ämnesklassning på 3-siffernivå vilket försvårar datauttag för ämnesanalyser. Inom vidareutveckling av Swepub ingår det att ta fram en prototyp för

¹² Nationella riktlinjer för beskrivning av vetenskaplig output 1.2. 2015, s. 20-21.

http://www.kb.se/dokument/SwePub/v-1.2-Nationella-riktlinjer-f%C3%B6r-beskrivning-av-vetenskaplig-output_2015_09_10.pdf

¹³ ORCID. <https://orcid.org/>

¹⁴ Approved changes for MODS 3.6. <http://www.loc.gov/standards/mods/changes-3-6.html>

¹⁵ MARC Code List for Relators. <http://www.loc.gov/marc/relators/relaterm.html>

semiautomatisk ämnesklassning. Genom data mining av abstract ska systemet föreslå ämnesklassning på 3-siffernivå i förhållande till liknande, ämnesklassificerade poster i Swepub.

Lärosätena uttrycker dock att SCB/UKÄ-klassningen inte motsvarar forskningsämnen fullt ut. Eftersom den internationella forskningsstatistiken utgår från OECD:s standard för ämnesklassning *Fields of Research and Development* (FORD) som SCB/UKÄ-klassningen baseras på, är det uteslutet att ändra standarden. Däremot kan det övervägas att införa kontrollerade nyckelord, till exempel utifrån *Svenska ämnesord*.

Metadatakrav för listor över forskningsoutput

Både VR och lärosätena har uttryckt behov att använda Swepub för import av listor över forskningsoutput. Att kunna sammanställa en lista över all forskningsoutput av en forskare utifrån Swepub skulle förenkla CV-hantering i både ansökningshanteringssystemet Prisma och forskarnas egna profilsidor. Följande uppgifter underlättar sammanställning och import av forskningsoutput enligt VR:s beställarunderlag:

- Upphovspersoner
- Titel
- Värdepublikation
- Innehållsmärkning
- Forskningsoutput
- OA typ
- ORCID
- ISI ID (UT-nummer)
- Länk till fulltext

Eftersom namn på upphovsperson registreras enligt riktlinjerna så som det förekommer i outputen¹⁶, försvårar det datauttag utifrån personnamn. Därmed är det viktigt med personID som ORCID. För angivning av ORCID i Swepub MODS har använts subelementet *description* men MODS har sedan dess uppdaterats med subelementet *nameIdentifier*. Identifierare ska anges som koder utan prefix och suffix men för att stödja länkade data borde det i namnrymden (namespace) anges http-URI för ORCID.

Innehållsmärkning beskriver huruvida outputen är sakkunniggranskad, övrig vetenskaplig eller övrig (populärvetenskap, debatt). Detta anses dock vara godtyckligt vilket är en av anledningarna varför VR vill införa en auktoriserad lista över publiceringskanaler som avgör om en värdepublikation är tillräckligt vetenskaplig eller inte. För avgränsning i sökverktyg, sammanställning av outputlistor och flaggning för nya publiceringskanaler som behöver bedömas av VR är det ändå användbart att ha innehållsmärkningen kvar. I så fall borde det undersökas om en justerad innehållsmärkning som endast avgör om en output är granskad eller inte motsvarar behoven bättre. Sådan innehållsmärkning skulle stödja även konstnärlig output som snarare är kollegialt granskad än sakkunniggranskad och nödvändigtvis inte har traditionella typer av publiceringskanaler. Indikation av den tänkta målgruppen som också har varit en av funktionerna för den nuvarande innehållsmärkningen kan ersättas med information om målgrupp.

Inom ramen för Swepub-projektet har det tagits fram nya outputtyper som bättre motsvarar behov av att beskriva även konstnärlig och konferensoutput i stället för de gamla publikationstyperna.¹⁷ Det har

¹⁶ Nationella riktlinjer för beskrivning av vetenskaplig output 1.2. 2015, s. 8.

http://www.kb.se/dokument/SwePub/v-1.2-Nationella-riktlinjer-f%C3%B6r-beskrivning-av-vetenskaplig-output_2015_09_10.pdf

¹⁷ Svensk indelning av vetenskaplig output 2015. http://www.kb.se/dokument/SwePub/Svensk-indelning-av-vetenskaplig-output_2015.pdf

dock kommit förändringsönskemål. Utifrån dessa önskemål kommer ett nytt förslag för revidering av outputtyperna att introduceras som också formatet behöver ta hänsyn till samt en reviderad mappningstabell mellan de nya outputtyperna och de gamla publikationstyperna för mappningen i Swepub och de lokala publikationsdatabaserna.

För att markera output som är parallellpublicerad med öppen tillgång (ÖT) i ett lärosätesarkiv (grön ÖT) har Swepub MODS utvecklats med attributet *note* i *URL*-elementet. Markering av output som har publicerats mot avgift i en prenumerationsbaserad tidskrift, det vill säga hybrider, och output med avgränsad tillgång samt information om licenser saknas. Markering av ÖT borde ses över i förhållande till MODS och e-pliktens MODS som använder sig av elementet *accessCondition* för information om tillgänglighet och licenser.

Swepub använder sig av datakällorna Directory of Open Access Journals (DOAJ), Directory of Open Access Books (DOAB) samt oaDOI för att markera guld ÖT det vill säga output publicerad i open access-tidskrift. Tjänsten oaDOI levererar även uppgiften om hybrider. Frågan är om data i Swepub behöver utvecklas med uppgift om hybrider när det går att berika data med denna uppgift från externa datakällor. Om det anses önskvärt att utveckla Swepub med uppgift om hybrider borde det undersökas om den går att generera genom uppgiften om tillgång och licens kopplad mot förlagsdata, till exempel via CrossRef, för information om förlagets publiceringslicenser.

För uppföljning av publiceringskostnader samlar KB:s databas Open APC Sweden data över författaravgifter (Article Processing Charge, det vill säga APC)¹⁸. I detta arbete vore det till hjälp om det fanns uppgift om corresponding author i Swepub. *MARC Code List for Relators* som används för roller i Swepub saknar dock denna roll. För kopplingen till Open APC Sweden skulle det också vara användbart med uppgift om APC i Swepub.

Kartläggning

För att kunna möta de behoven som har beskrivits i detta dokument har följande format utöver MODS undersökts: Dublin Core, CERIF och BIBFRAME.

BIBFRAME

BIBFRAME¹⁹ utvecklas av Library of Congress och är ämnat att ersätta MARC som format för bibliografisk information som länkade data. Tanken är att BIBFRAME inte enbart ska rikta sig mot biblioteksverksamheter. BIBFRAME är fortfarande under utveckling och har inte börjat användas i stor omfattning. Libris XL är bland de första som implementerar BIBFRAME, i en modifierad version. På grund av den pågående utvecklingen kommer formatet inte att kunna användas för Swepubs behov i första hand. Dock har det som format som stödjer bibliografiska, länkade data mycket potential att utvecklas till framtidens format även för Swepubs räkning. BIBFRAME liknar den bibliometriska datamodellen i Swepub där flera instanser det vill säga poster från olika organisationer pekar åt ett verk, samarbetspublikationen, men behöver anpassas för andra bibliometriska funktionaliteter. Eftersom Libris XL använder BIBFRAME som grund i sitt format finns det både kunskap och synnergieffekter. För att kunna dra större nytta av data och infrastruktur mellan Swepub och Libris kommer det bli aktuellt att mappa Swepub MODS mot BIBFRAME. Det finns redan ett förslag på mappning mellan MODS 3.6 och BIBFRAME 2.0 av Library of Congress som kan användas för den interna mappningen mellan Swepub och Libris.²⁰

Fördel med BIBFRAME	Nackdel med BIBFRAME
Synnergieffekten med Libris XL	Under utveckling
Kunskap inom KB	Ämnat för bibliografiska metadata

¹⁸ Open APC Sweden. <https://github.com/Kungbib/openapc-se>

¹⁹ BIBFRAME. <http://www.loc.gov/bibframe/docs/index.html>

²⁰ MODS to BIBFRAME Mapping. <https://www.loc.gov/standards/mods/modsrdf/mods-bibframe-mapping.html>

Utvecklingsbart	Inte etablerat
	Kräver mappningar mot andra format för att data blir användbara utanför KB

CERIF

CERIF (Common European Research Information Format)²¹ är som namnet antyder framtagen för beskrivningsbehoven av forskningsinformation och används i CRIS-system (Current Research Information Systems). CERIF kan användas för beskrivning av forskningsentiteterna, deras aktiviteter och output samt relationer. Det möjliggör åtkomst av forskningsinformation och informationsbyten mellan forskare, förlag, finansiärer, forskningsutvärdering och andra aktörer inom forskningsinformation samt publik. Därmed motsvarar det väldigt bra bibliometriska analysändamål men har inte bibliografiska beskrivningsbehov som utgångspunkt. CERIF har omfattande syntax för entiteter inom forskningsinformation. Samtidigt är det komplext och därmed tungrott att implementera. Utanför CRIS-system är det dessutom relativt oetablerat format vilket skulle kunna vara ett hinder för Swepubs dataleverantörer och användare. Det finns heller ingen kunskap och därmed inga synnergieffekter inom KB eftersom formatet inte används i KB-system sedan tidigare.

Fördel med CERIF	Nackdel med CERIF
Ämnat för forskningsinformation i stort	Inte tillräckligt uttrycksfull för bibliografiska metadata
Mycket användbart för bibliometriska analysändamål	Ingen kunskap och synnergieffekt inom KB
Omfattande syntax	Komplext
Synnergieffekter mellan CRIS-system	Inte etablerat utanför forskningsinfrastruktur
	Kräver mappningar mot andra format för att data blir användbara utanför KB

Dublin Core

Dublin Core (DC) är ett välkänt format och underhålls av Dublin Core Metadata Initiative (DCMI), öppen, non-profit organisation, som aktivt arbetar för att främja metadatabeskrivningar.²² DC används till exempel i Finland för beskrivning av forskningsoutput²³. DC består ursprungligen av 15 element, numera kända som Dublin Core Metadata Element Set (DCMES)²⁴ samt ytterligare tillagda termer som tillsammans bildar DC Metadata Terms – vokabulären. Det finns även anpassning för bibliotek, DC Library Application Profile²⁵. DC används brett i alla typer av system vilket förenklar återanvändning av data. Exempelvis används det av DOAJ samt för e-pliktens RSS-flöden. DC har fördelen att vara väl utbredd och allsidigt format vilket är också dess nackdel. Eftersom det är anpassat för alla typer av verksamheter kan det vara svårt att få till all den funktionaliteten som bibliometriska analysändamål sätter krav på. Enligt Finlands metadatarekommendation för textpublikationer i publikationsdatabaser togs inte DC Metadata Terms i bruk så som de rekommenderas av DCMI eftersom dessa inte skulle ha möjliggjort tillräckligt uttrycksfull och enhetlig beskrivning av metadata.²⁶ Dessutom gick det inte i Finland att hitta en tillfredsställande lösning för affilieringar. Detta skulle kräva liknande vidareutveckling som har skett för MODS.

²¹ CERIF. <https://www.eurocris.org/cerif/main-features-cerif>

²² Dublin Core Metadata Initiative. <http://dublincore.org/about/>

²³ Publikationsarkivtjänsterna. <https://www.kiwi.fi/pages/viewpage.action?pageId=44237266>

²⁴ Dublin Core Metadata Element Set. <http://dublincore.org/documents/dces/>

²⁵ DC Library Application Profile. <http://dublincore.org/documents/library-application-profile/>

²⁶ Metadatausuositus julkaisuarkistojen tekstijulkaisuille (på finska).

<https://www.kiwi.fi/display/Julkaisuarkistopalvelut/Metadatausuositus+julkaisuarkistojen+tekstiaineistolle>

Fördel med DC	Nackdel med DC
Välkänt	Saknar bibliometriska funktioner
Allsidigt	Inte tillräckligt uttrycksfull för bibliografiska metadata
Kunskap inom KB	Begränsade synenergieffekter inom KB
Synenergieffekter mellan olika typer av system utanför KB	

MODS

Swepub använder sig av en modifierad version av MODS. För återanvändning av data är det dock viktigt att Swepub MODS inte utvecklas alldeles för mycket avvikande från det ursprungliga MODS. I saknaden av viss funktionalitet i MODS har Swepub vidareutvecklat det för att få formatet att motsvara de bibliometriska behoven. Sedermera har en del av de saknade funktionaliteterna införts i MODS, till exempel *nameIdentifier* för personID. Med tanke på att MODS utvecklas mot den inriktning som Swepub är i behov av vore det rimligt att fortsätta använda MODS som överföringsformat i Swepub. Dessutom används MODS även i e-plikten vilket bidrar till synenergieffekter och kunskap inom KB. Med tanke på att det finns användare av Swepub-data som inte är lika insatta i terminologin i Swepub MODS behöver vokabulär dock förtydligas och publiceras.

Nedan listas de element som behöver revideras, införs och/eller avlägsnas i Swepub MODS för att motsvara de bibliometriska behoven och följa MODS:

- Införa elementet *accessCondition* med värden "restricted" för output med avgränsad tillgång och "gratis" för output med öppen tillgång i stället för att ange denna information i elementet *location* med *note="free"*
- Införa attributet *type="use and reproduction"* för *accessCondition* för information om licenser och embargo
- Införa subelementet *nameIdentifier="orcid"* i stället för att använda subelementet *description* för det
- Införa attributet *invalid* för att markera identifikatorer som ogiltiga
- Byta ut attributet *authority* i elementet *genre* från *sv* till *kb.se* och införa en ny värdelista för innehållsmärkningen bestående av "reviewed" och "non-reviewed"
- Införa attributet *authority* i elementet *subject* för att ge möjlighet att ange auktoriserade ämnesord utifrån *Svenska ämnesord* och *Library of Congress Subject Headings*

En ny version av MODS, 3.7,²⁷ har precis släppts och innehåller en del av funktionaliteten som efterfrågas i Swepub MODS. Möjlighet att ange auktoriteter för förlag har införts vilket kommer lämpligt med tanke på den nordiska listan och att KB har blivit medlem i ISNI (International Standard Name Identifier). Kontrollerad lista på *typeOfResource* har avlägsnats vilket kommer att underlätta beskrivning av konstnärlig output. Även namnvarianter har införts vilket kan vara användbart i Swepub.

Fördel med MODS	Nackdel med MODS
Utvecklas kontinuerligt	Ämnat i grunden för bibliografiska metadata
Har redan vidareutvecklats för bibliometriska analysändamål inom Swepub-projektet	Finns begränsningar på hur mycket det går att avvika från det ursprungliga MODS
Kunskap inom KB	Inte lika etablerat utanför kulturarvsområdet

²⁷ MODS 3.7 announcement. <http://www.loc.gov/standards/mods/mods-3-7-announcement.html>

Synenergieffekter med e-plikten	
Mapperingar mot MARC och DC samt en under arbete mot BIBFRAME	

Rekommendation

För att kunna uppfylla de metadatakrav som Swepubs främsta intressenter, VR och SUHF, har, krävs det ett format som både har avancerade bibliografiska och bibliometriska funktionaliteter. I dagsläget motsvarar inget av de kartlagda formaten både de bibliografiska och de bibliometriska beskrivningsbehoven inom Swepub fullt ut.

Ett formatbyte är omfattande arbete och inte enbart inom KB. Formatbyte skulle innebära förlängning på projektet men även mycket arbete för dataleverantörerna att implementera ett nytt format. I det rådande läget där dataleverantörerna vid lärosätesbiblioteken redan utsetts för både format- och praxisbyte inom katalogisering samtidigt som alla myndigheter har krav på sig att leverera e-plikt skulle det kunna vara svårt att motivera ett formatbyte utan att kunna visa betydlig nytta med det. Eftersom det är frivilligt att leverera data till Swepub finns det heller inget incitament för KB att kräva ett sådant byte från de deltagande organisationerna.

Av de format som har undersökts motsvarar CERIF bäst beskrivningsbehoven av forskningsinformation men utifrån alla forskningsaktiviteter så som projekt, i stället för att utgå från att beskriva forskningsoutput. Därmed lämpar det sig bättre till fullvärdiga CRIS-system än databaser ämnade för publikationer. Det finns heller inga synenergieffekter då formatet inte är etablerat utöver CRIS-system.

Dublin Core är ett enkelt format för dataleveranser då det är ett välanvänt format i många system vilket skulle underlätta informationsbyte mellan system. Dock är det inte tillräckligt uttrycksfull för att beskriva forskningsoutput så som Swepub behöver.

BIBFRAME är fortfarande under utveckling vilket kan bidra till stora utvecklingsmöjligheter samtidigt som det behöver utvecklas mer för att kunna användas som överföringsformat för Swepub för både bibliografiska och bibliometriska behov. Däremot kommer det att behövas en mappning mellan Libris BIBFRAME och Swepub MODS för att utnyttja de synenergieffekter som finns mellan Libris och Swepub. När tiden är inne och funktionaliteten finns i BIBFRAME kan det även bli aktuellt att byta Swepub MODS som överföringsformat mot BIBFRAME. Detta kommer dock inte att vara möjligt under den närmaste framtiden utan kommer att kräva en längre utveckling.

Eftersom MODS redan har anpassats för Swepub och utvecklingen av MODS sker mot den riktning som motsvarar behoven och önskemålen hos Swepubs intressenter, rekommenderas att Swepub MODS behålls som överföringsformat för Swepub och att det vidareutvecklas enligt de senaste versionerna av MODS.