



Vetenskapsrådet

Datum
2017-09-21

Diarienummer
3.1-2017-6129

Handläggare
Henrik Aldberg

Beställarunderlag för ett analyserbart SwePub

Kungliga biblioteket (KB) har i uppdrag att utveckla databasen SwePub så att den går att använda till bibliometriska analyser [1]. Vetenskapsrådet (VR) har i uppdrag att samverka med KB i detta arbete [2]. Arbetet med (och därmed kostnaden för) att förbättra kvaliteten på data ligger till stor del hos de organisationer som levererar data till SwePub. Detta dokument beskriver vilka datafält i SwePub som måste hålla god kvalitet för att SwePub ska kunna användas till

1. Bibliometriska analyser
2. Import av publikationer till Prisma.

Dokumentet är tänkt att användas som ett underlag för de organisationer som levererar data till SwePub då de ska bedöma kostnaden för att förbättra datakvaliteten.

I dagsläget är det 41 organisationer som levererar data till SwePub. Informationen publiceras sedan på två olika platser: i söktjänsten *SwePub* som innehåller bibliografisk information över vetenskaplig publicering vid svenska universitet och högskolor samt forskande myndigheter, och i *SwePub för analys och bibliometri* som är en tjänst för att ta fram bibliometrisk statistik. Om olika organisationer redovisar samma publikation kombineras informationen innan den visas i söktjänsten medan analystjänsten visar samtliga redovisade publikationer med en flagga som markerar dubletter.

En förutsättning för ett analyserbart SwePub är att det finns en auktoriserad svensk lista över publiceringskanaler som kan kopplas till SwePub via ISSN eller ISBN. En sådan lista är viktig för att på ett automatiserat sätt avgöra om en publiceringskanal är vetenskaplig. KB och VR ingår i ett nordiskt samarbete som syftar till att slå samman de auktoriserade listor som finns i Norge, Finland och Danmark vilket kommer att förenkla arbetet med att bedöma publiceringskanalerna i SwePub.

I dokumentet *Nationella riktlinjer för beskrivning av vetenskaplig output – Version 1.2* beskrivs de nationella riktlinjer som finns för hur innehåll ska levereras till SwePub [3]. Alla fält som listas nedan är obligatoriska enligt dessa riktlinjer, utom fältet *Totalt antal lokala upphovspersoner uppdelat på kön* i Tabell 1.

Dokumentet *Rekommendationer och förslag på åtgärder för vidareutveckling av SwePub* presenterar de rekommendationer och förslag på åtgärder för vidareutveckling av SwePub som tagits fram av SwePub-projektet [4]. Dessa rekommendationer är indelade i tio områden¹ och kolumnen Område i tabellerna nedan anger vilket område respektive datafält hamnar under.

¹ De tio områdena är: 1. Övergripande vidareutveckling av SwePub, 2. Format, 3. Upphov, 4. Identifierare, 5. Uppgifter om utgivning, 6. Ämne, 7. Outputtyper, 8. Tillgång, 9. Medietyp och 10. Innehållsmärkning.

1 – Bibliometriska analyser

De fält som listas nedan behövs för att koppla publiceringskanaler till den nordiska listan samt göra sammanställningar över antalet unika publikationer per år och svensk organisation. Tanken är att när projekt rapporteras i SweCRIS används DOI för att få en koppling mellan publikationerna i SwePub och projekten i SweCRIS.

Fälten i detta område kan delas in i två olika kvalitetsnivåer där Tabell 1 innehåller fälten på den första nivån och Tabell 2 innehåller fältet på den andra nivån.

För de fält som innehåller en identifierare, till exempel ISSN, DOI och ORCID, är det mycket viktigt att de lokala systemen utformas så att det inte går att mata in data på ett felaktigt format. I dagsläget är det vanligt att ISSN för den tryckta versionen och ISSN för den elektroniska versionen registreras i samma fält.

Fält	Kommentar	Område (enligt [4])
ISSN (ISBN)	För en koppling mot den nordiska listan över auktoriserade publikationskanaler samt Directory of Open Access Journals (DOAJ) och liknande listor.	4
Publiceringsstatus	Opublicerat material ska märkas med någon av de fem kategorierna: Preprint, Submitted, Accepted, In press och Epub ahead of print/Online first. Material som saknar publiceringsstatus räknas som Published. ²	5
Publiceringsår	Behövs för att kunna studera olika tidsperioder.	5
Digital Object Identifier (DOI) (Unik SwePub-identifierare)	Det behövs någon form av unik identifierare som är beständig över tid. DOI vore det bästa men en SwePub-specifik identifierare kan vara en tillfällig lösning.	1
Affilieringar	I [3] är uppgift om affiliering med organisationsidentifierare obligatorisk för den egna organisationens upphovspersoner. För externa upphovspersoner från Sverige är affiliering frivilligt då det fås från andra lärosäten. För externa upphovspersoner med affiliering till utländskt lärosäte skall land uppges som affiliering för upp till tio personer.	3
Totalt antal upphovspersoner	Behövs för att kunna författarfraktionera en publikation.	3
Totalt antal lokala upphovspersoner uppdelat på kön	I [3] är det obligatoriskt att registrera alla upphovspersoner från det egna lärosätet men det är inte obligatoriskt att registrera kön.	3

Tabell 1 - Fält på kvalitetsnivå 1.

² Det kan förekomma Epub ahead of print/Online first som publiceras långt före den tryckta utgåvan och hinner bli citerad innan den tryckta versionen kommer. Epub ahead of print/Online first tilldelas ofta DOI för att kunna bli citerade och sedan november 2016 har CrossRef börjat med tilldelning av DOI för preprints [4, sid. 12].

Det enda fält som listas i Tabell 2 är forskningsämne utifrån SCB:s 3-siffernivå. Detta tillhör egentligen minimikraven för bibliometriska analyser men då detta medför mycket arbete för lärosätena har det lagts på en separat nivå. En lösning kan vara att alla publikationer klassas maskinellt av SwePub³ men om ett lärosäte valt att ämnesklassa en publikation så används denna klassning.

Fält	Kommentar	Område (enligt [4])
Forskningsämne utifrån SCB:s 3-siffernivå	Ska detta göras manuellt krävs det mycket arbete från lärosätena. Ett alternativ på sikt är att lärosätena inte klassar publikationerna själva och SwePub använder en automatisk ämnesklassning för samtliga publikationer.	6

Tabell 2 - Fält på kvalitetsnivå 2.

2 – Användning i Prisma

De fält som är relevanta om SwePub ska användas för att importera publikationer till Prisma listas i Tabell 3. Några avfälten här är primärt avsedda för manuell granskning vid hantering av ansökningar snarare än bibliometriska analyser. Till exempel är det för en granskare inte lika viktigt att personnamn i fältet Upphovspersoner är formaterat på ett standardiserat sätt medan fältet ORCID måste ha en strikt standard för hur det fylls i då det är detta fält snarare än personnamn som används för koppling mot andra datakällor.

Fält	Kommentar	Område (enligt [4])
Upphovspersoner	I [3] är minimikravet att namn på den första upphovspersonen, samtliga upphovspersoner från det egna lärosätet samt sista upphovspersonen anges i den ordning de står på publikationen. Om inte corresponding author finns bland dessa ska denne också uppges.	3
Titel		-
Värdpublikation		5
Innehållsmärkning	Det finns tre nivåer: Sakkunniggranskat, Övrigt vetenskapligt och Övrigt. Den auktoriserade listan är tänkt att bestämma vad som är sakkunniggranskat.	10
Forskningsoutput	Det finns fem huvudkategorier: Konstnärlig output, publikationer, Konferensoutput, Immaterialrättslig output och Övrig output (för dataset).	10
OA typ	Med ISSN går det att kontrollera om en tidskrift finns i DOAJ och SwePub kompletterar med metadata från tjänsten oaDOI. För att användas för uppföljning behöver dock SwePubs formatmall ses över.	8
ORCID	Om lärosätet använder ORCID är det enligt [3] obligatoriskt att leverera detta. Är även viktigt för bibliometriska analyser.	3
ISI ID (UT-nummer)	Denna identifierare är prioriterad då VR i nuläget har en lokal version av WoS. KB och den nationella styrgruppen för forskningsinformation (NSFI) har i uppdrag att utreda om det är tillåtet att göra denna identifierare öppet tillgänglig i SwePub.	4
Länk till fulltext	Intressant för uppföljning av projekt.	8

Tabell 3 - Fält som är relevanta för import av publikationer till Prisma.

³ Till exempel har Linköpings universitet utvecklat ett system som utifrån titel och abstract automatiskt ämnesklassar publikationer.

3 – Övriga dokumentidentifierare

Dessa fält gör det enklare att göra kopplingar till internationella publikationsdatabaser. Det behöver undersökas om det är möjligt att sprida dessa identifierare utan att bryta mot respektive kontrakt (gäller även ISI ID som nämns ovan).

Fält	Kommentar	Område (enligt [4])
Scopus ID	Unik identifierare som används i Scopus.	4
Pubmed ID	Unik identifierare som används i Pubmed.	4

Relaterade dokument

[1] Regleringsbrev för budgetåret 2017 avseende Kungl. biblioteket

[2] Regleringsbrev för budgetåret 2017 avseende Vetenskapsrådet

[3] Nationella riktlinjer för beskrivning av vetenskaplig output – Version 1.2, KB dnr. 1.4.1-2015-824

[4] Rekommendationer och förslag på åtgärder för vidareutveckling av SwePub, KB dnr. 6.4-2016-337

[5] Svensk indelning av vetenskaplig output 2015

[6] Gruppen för SwePub på webben