

SwePubs lathund för databearbetning

Version 1.0

Dnr 1.4.1-2016-776



National Library
of Sweden

Innehållsförteckning

| | | |
|-----------|--|-----------|
| 1. | INLEDNING | 3 |
| 1.1. | Om lathunden | 3 |
| 1.2. | Kontroll och validering av metadata | 3 |
| 1.3. | Databearbetning | 4 |
| 1.4. | Flaggning av icke-validerade poster..... | 5 |
| 1.5. | Arbetsgrupp och kontakt | 5 |
| 2. | KONTROLL OCH VALIDERING AV METADATA . | 6 |
| 2.1. | Allmänt | 6 |
| 2.2. | För- och efternamn/Organisation som upphov | 6 |
| 2.3. | Affilieringar..... | 7 |
| 2.4. | PersonID | 8 |
| 2.5. | Roll och antal upphovspersoner (creator count).... | 9 |
| 2.6. | Titel..... | 9 |
| 2.7. | Utgivningsdatum | 9 |
| 2.8. | Utgivare | 10 |
| 2.9. | Värdpublikationens titel..... | 10 |
| 2.10. | Output..... | 11 |
| 2.11. | Identifikatorer för output och publiceringskanaler | 11 |
| 2.11.1. | DOI | 12 |
| 2.11.2. | ISI ID..... | 12 |
| 2.11.3. | ISBN | 12 |
| 2.11.4. | PubMed ID..... | 12 |
| 2.11.5. | ISSN | 12 |
| 2.12. | Publiceringsstatus..... | 12 |
| 2.13. | Åtkomst till fulltext | 13 |
| 2.14. | Ämne | 14 |
| 2.15. | Innehållsmärkning..... | 15 |
| 3. | HANTERING AV OFULLSTÄNDIG METADATA. | 16 |
| 3.1. | Utsökning av feltyper | 16 |
| 3.2. | Export av feltyper..... | 16 |
| 3.3. | SwePub:s feltyper..... | 17 |
| 3.3.1. | DOI fel | 17 |
| 3.3.2. | ISBN på fel nivå | 18 |
| 3.3.3. | ISI ID fel | 18 |
| 3.3.4. | ISSN fel | 18 |
| 3.3.5. | ISSN saknas | 19 |
| 3.3.6. | Konferens – saknad värdpublikation | 19 |

| | | |
|-----------|---|-----------|
| 3.3.7. | ORCID fel | 19 |
| 3.3.8. | PersonID fel | 19 |
| 3.3.9. | PersonID saknas..... | 20 |
| 3.3.10. | Upphov – antal saknas | 20 |
| 3.3.11. | Upphov – olika antal | 20 |
| 3.3.12. | Upphov saknas | 21 |
| 3.3.13. | Utgivningsdatum fel | 21 |
| 3.3.14. | ISBN – fel landkod | 21 |
| 3.3.15. | ISBN fel | 22 |
| 3.3.16. | Publiceringsstatus fel | 22 |
| 3.3.17. | Upphov – dubblering..... | 22 |
| 3.3.18. | Upphov – olika namnvarianter..... | 23 |
| 3.3.19. | Ämne saknas | 23 |
| 4. | HANTERING AV DUBBLETTKANDIDATER..... | 24 |
| 4.1. | Utsökning av dubblettkandidater | 24 |
| 4.2. | Bedömning av dubblettkandidater..... | 25 |
| 4.3. | Export av dubblettkandidater | 25 |
| 5. | HANTERING AV SAMARBETSPUBLIKATIONER | 26 |
| 5.1. | Utsökning av samarbetspublikationer | 26 |
| 5.2. | Bedömning av samarbetspublikationer | 27 |
| 5.3. | Export av samarbetspublikationer | 27 |
| 5.4. | Motstridigheter | 28 |
| 5.4.1. | Fraktionering stämmer ej | 29 |
| 5.4.2. | Antal upphov olika..... | 29 |
| 5.4.3. | DOI olika | 29 |
| 5.4.4. | Kanaler olika | 30 |
| 5.4.5. | År olika..... | 30 |
| 5.4.6. | Outputtyper olika..... | 30 |
| 5.4.7. | Ämnen olika | 31 |
| 5.4.8. | Titlar olika | 31 |
| 6. | APPENDIX..... | 32 |
| 6.1. | Kravspecifikation för metadata..... | 32 |
| 6.2. | Valideringskällor | 34 |
| 6.3. | Tips på källor för maskinell validering | 36 |
| 6.3.1. | Validering av identifikatorer | 36 |
| 6.3.2. | Validering av open access | 36 |

1. INLEDNING

1.1. Om lathunden

Det övergripande syftet med lathunden är att stödja datakvalitetshöjande arbete vid deltagande organisationer genom att förtydliga de krav som ställs på metadata för att kunna ligga till grund för bibliometriska analyser. Lathunden är tänkt att ge stöd och tips för validering samt ge anvisningar för databearbetningen i tjänsten SwePub för analys och bibliometri.¹

Målgruppen för dokumentet är personer som registrerar och kvalitetssäkrar output i lokala publikationsdatabaser som deltar i och levererar data till SwePub. Lathunden bör användas som ett komplement till *Nationella riktlinjer för beskrivning av vetenskaplig output 1.2 (NR)*, formatspecifikationen för *SwePub MODS 2.6 (MODS)* och *Svensk indelning av vetenskaplig output 2015 (SIVO)*².

1.2. Kontroll och validering av metadata

Kvalitetssäkrad och validerad³ metadata är en nyckel till tillförlitlig forskningsinformation. Fullständig metadata underlättar identifiering, länkning, åtkomst och återanvändning av forskningsinformation. Den underlättar också uppföljning och analyser av olika slag samt nationella och internationella jämförelser.

Lathunden är primärt utarbetad för att stödja det manuella arbetet med kvalitetssäkring av metadata. Den utgår ifrån kravspecifikationen för metadata för leverans till SwePub. Kravspecifikationen ger instruktioner om vilka fält som är obligatoriska för att möjliggöra analyser och uppföljningar. Dessa fält har mappats mot fält i *MODS* och *NR* för mer utförlig information och instruktioner för registrering.⁴

I syfte att minska mängden av poster för manuella åtgärder rekommenderas dock att först ta reda på omfattningen av arbetet, vilka data som går att rätta till maskinellt samt vad som eventuellt går att integrera maskinellt i lokala system. Genom att utveckla inmatningsformulär i lokala publikationsdatabaser och sätta regler för den manuella inmatningen av data i gränssnittet går det att minska datakvalitetsbrister, till exempel genom att ta bort möjligheten att ange identifikatorer som DOI med ogiltiga tecken och på fel nivå. Ett annat sätt att minska datakvalitetsbrister är att använda externa källor för verifiering av metadata. Tips på maskinell

¹ SwePub för analys och bibliometri: <http://bibliometri.swepub.kb.se>

² Dokumenten finns samlade i en verktygslåda för beskrivning av vetenskaplig output: <http://www.kb.se/libris/SwePub/Format-och-praxis/>

³ Med validering menas att bekräfta giltighet, att officiellt godkänna ett dokument och försäkra att det som står där är korrekt.

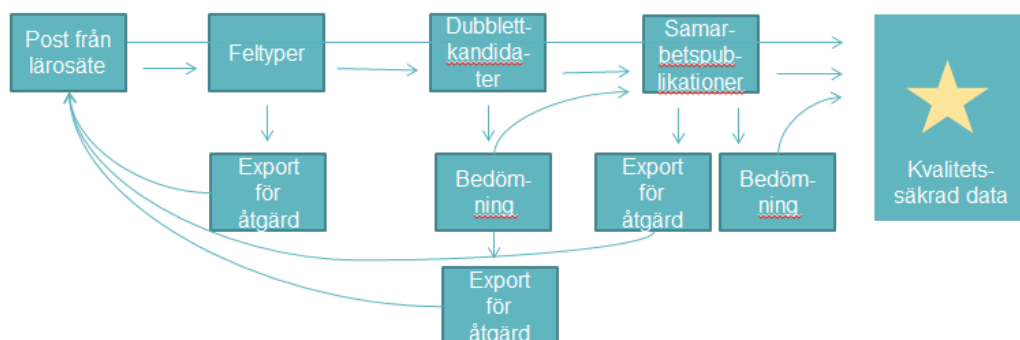
⁴ Appendix 6.1. Kravspecifikation för metadata

verifiering finns i appendixet.⁵ Vid manuella insatser är ett gott råd att åtgärda alla eventuella brister för en enskild post vid samma tillfälle.

1.3. Databearbetning

Tjänsten SwePub för analys och bibliometri är framtagen för att stödja uttag av data om vetenskaplig output för analys och databearbetning.⁶ En bra utgångspunkt i arbetet med kvalitetssäkring är att först åtgärda den output som är publicerad inom tidsramen för möjliga analyser. En annan är att avgränsa till de outputtyper som räknas med i analyser, till exempel artiklar i vetenskapliga tidskrifter och paper i proceeding.

Det finns tre steg mot högre datakvalitet: hantering av ofullständig metadata (SwePubs feltyper), dubblettkandidater och motstridigheter i samarbetspublikationer. Målet är kvalitetssäkrad data som kan ligga till grund för bibliometriska analyser.



Kvalitetssäkringsprocessen i SwePub

Hantering av ofullständig metadata är avgörande för att poster ska gå vidare i processen till dubbletthanteringen. Om det finns brister i datakvaliteten märker systemet upp poster med så kallade feltyper efter regelkörningen i SwePub. Dessa poster behöver då åtgärdas lokalt. Prioritera de åtgärder som har störst påverkan på processen för kvalitetssäkrad metadata och därmed inkludering i analysunderlag. Poster som har brister i metadata av hög bibliografisk karaktär är märkta med allvarlighetsgrad 3, vilket exkluderar dem från dubbletthanteringen och analysunderlag. Utgå därför i första hand från dessa. Poster med feltyper av allvarlighetsgrad 1-2 påverkar vissa typer av analyser men inkluderas i dubbletthanteringen.

⁵ Appendix 6.3. Tips på källor för maskinell validering

⁶ SwePub för analys och bibliometri: <http://bibliometri.swepub.kb.se/>

Hantering av dubblettkandidater är avgörande för att poster ska inkluderas i analysunderlag eftersom systemet inte klarar av att maskinellt bedöma dem som dubletter eller icke-dubletter. Poster som märks upp som dubblettkandidater behöver bedömas i SwePub och eventuellt åtgärdas lokalt.

Hantering av samarbetspublikationer är det tredje och sista steget på väg mot kvalitetssäkrad data. Poster som har blivit bedömda som dubletter går igenom en regelkörning för uppmärkning av motstridigheter. Dessa poster behöver hanteras lokalt. Poster som felaktigt har identifierats som dubletter av systemet behöver bedömas i SwePub för att överskrida den maskinella bedömningen.

1.4. Flaggnings av icke-validerade poster

Ett lärosäte kan välja att flagga nyregistrerade poster som inte har genomgått en lokal granskning. Poster med flaggning utesluts från kvalitetssäkringsprocessen i SwePub och ingår endast i uttag av originaldata. Efter slutförd granskning plockar lärosätet bort flaggningen. Poster utan flagga betraktas som granskade och ingår i kvalitetssäkringsprocessen.

1.5. Arbetsgrupp och kontakt

Lathunden är utarbetad av projektgruppen för Vidareutveckling av Swepub, Kungliga biblioteket.

Kommentarer och frågor om lathunden kan skickas till: libris@kb.se

2. KONTROLL OCH VALIDERING AV METADATA

För att höja datakvaliteten och därmed säkra att data räknas med i uttagen för bibliometriska analyser bör följande data kontrolleras och valideras. De anges här tillsammans med förslag på åtgärder för databearbetning med hänvisning till ”Kravspecifikation för metadata” i de fall datafältet finns listat där.⁷ I appendix listas de valideringskällor som SwePub använder samt tips för källor för maskinell validering.⁸

2.1. Allmänt

- Om tecken inte är registrerade enligt teckenkodning UTF-8 kommer även felfri data att betraktas som felaktig. Tänk på detta även vid kopiering och import av data.
- Identifikatorer registreras enbart som maskinläsbar kod, med bindestreck men utan prefix som ”WoS” eller ”PDF”.
- Identifikatorer, affilieringar och ämnesklassifikationer bör registreras var för sig i egna fält, till exempel ISBN för den tryckta och den elektroniska publikationen.
- Identifikator för värdpublikation, framför allt ISBN och DOI, anges inte som identifikator för enstaka kapitel.
- Kontrollera för felskrivningar, att data inte ligger i fel fält, till exempel ISBN i fältet för ISSN, eller att fält inte har fyllts med frågetecken eller streck vid saknade uppgifter.
- Output kan ha blivit tilldelad olika identifikatorer vid olika skeden i publiceringen. Kontrollera att identifikatorerna gäller för den publicerade versionen.

2.2. För- och efternamn/Organisation som upphov

Det är viktigt att registrera alla upphovspersoner eftersom antalet upphovspersoner vid fraktionering räknas utifrån angivna namn. I de fall där antalet upphovspersoner är mycket omfattande och därmed arbetskrävande, kan upphovspersonerna anges enligt minimikravet beskrivet i NR: första upphovspersonen, alla egna upphovspersoner samt den sista upphovspersonen tillsammans med manuellt justerad uppgift om det egentliga antalet upphovspersoner.⁹

| Datafält | MODS | NR |
|---------------------|---|----------|
| <i>Person name,</i> | Mandatory for originators affiliated with the organisation | Se nedan |

⁷ Appendix 6.1. Kravspecifikation för metadata

⁸ Appendix 6.2. Valideringskällor och 4.3 Tips på källor för maskinell validering

⁹ NR, avsnitt 2.1 För- och efternamn & avsnitt 2.7 Organisation som upphov:
http://www.kb.se/dokument/SwePub/v-1.2-Nationella-riktlinjer-f%C3%B6r-beskrivning-av-vetenskaplig-output_2015_09_10.pdf

| | | |
|---|---|---|
| <i>identifier and affiliation</i> ¹⁰ | responsible for entering the record: <i>Names and affiliations</i> (13): All originators related to the organisation registering the record must be represented by a name element each. | |
| - <i>Name and affiliation</i> | Mandatory for originator persons related to the organisation responsible for entering the record: <i>Person name, identifier and affiliation</i> (13.1) | Se nedan |
| - <i>Name part</i> | Mandatory: <i>Name Part</i> (13.1.1) | Obligatoriskt: <i>För- och efternamn</i> (2.1): Alla upphovspersoner registreras så långt som möjligt, annars enligt minimikravet (första, sista och lärosätets egna upphovspersoner) |
| - <i>Organisation</i> | Mandatory when applicable: <i>Organisation as author</i> (13.2) Mandatory for names related to the organisation responsible for entering the publication record: <i>Affiliation</i> (13.2.1) | Obligatoriskt så som beskrivet under Upphovspersoner: <i>Organisation som upphov</i> (2.7) |

2.3. Affilieringar

Kontroll av affiliering är av stor vikt eftersom det påverkar fraktioneringen. En publikation ska alltid kopplas till en organisation genom att affiliering anges för den registrerande organisationens upphovspersoner. För att underlätta samarbetsanalyser ska affiliering anges på organisationsnivå för den registrerande organisationens upphovspersoner tillsammans med land för upp till tio externa upphovspersoner från organisationer utanför Sverige¹¹. För att möjliggöra djupare samarbetsanalyser kan även fakultets- och institutionsnivå anges.¹²

Felaktiga affilieringar kan upptäckas genom att göra uttag av samarbetspublikationer i SwePub för analys och bibliometri, till exempel när forskaren har angett sin nuvarande organisation som affiliering fastän forskaren var affilierad till en tidigare organisation vid

¹⁰ *Name part & Organisation* i Appendix 6.1. Kravspecifikation för metadata

¹¹ Detta gäller för output som inte ingår i Web of Science.

¹² För att stödja registrering av affiliering har en prototyp för auktoritetsregistret över organisationer upprättats. På sikt kommer prototypen vidareutvecklas och utökas med både svenska och internationella organisationer och centrumbildningar samt med landsregister. För mer information se: <https://github.com/libris/swepub-sparql/wiki/Auktoritetsregister-%C3%B6ver-publiceringskanaler-och-organisationer>

publiceringstillfället.¹³ Uttag av samarbetspublikationer kan även utnyttjas för att hitta saknade dubbelaffilieringar.

| Datafält | MODS | NR |
|------------------------------------|---|--|
| - <i>Affiliation</i> ¹⁴ | Mandatory for persons related to the organisation responsible for entering the publication record: <i>Affiliations</i> (13.1.3) | Obligatoriskt för den egna organisationens upphovspersoner. Frivilligt för externa upphovspersoner från Sverige: <i>Affilieringar</i> (2.3) |
| - <i>Country</i> | Mandatory for up to 10 external originators associated with organisations from countries other than Sweden: <i>Affiliation, country</i> (13.1.5) | Obligatoriskt för externa upphovspersoner med affiliering till ett utländskt lärosäte, för upp till 10 personer – för publikationer utanför Web of Science: <i>Affilieringar</i> (2.3), <i>Länder</i> (2.3.4) |

2.4. PersonID

För identifiering av upphovspersoner anges både lärosäteslokalt personID och/eller ORCID för lärosätets egna forskare i de fall de finns. ORCID är en nyckelkomponent för återanvändning av data inom ekosystemet för forskningsinformation. Ifall personID saknas, lämnas fältet tomt.

| Datafält | MODS | NR |
|--|--|---|
| - <i>Person identifier</i> ¹⁵ | Mandatory for originator persons related to the organisation responsible for entering the record: <i>Person name, identifier and affiliation</i> (13.1) | Obligatoriskt för lärosätets egna upphovspersoner: lärosäteslokalt <i>PersonID</i> (2.4.1) |
| - <i>ORCID</i> | Mandatory if applicable. Use for originators with a relation to the organisation entering the record, if the person has an ORCID: <i>Person identifier ORCID</i> (13.1.2) | Obligatoriskt för lärosätets egna upphovspersoner: <i>ORCID</i> levereras i de fall lärosätet använder dessa (2.4.2) |

¹³ SwePub för analys och bibliometri/databearbetning: <http://bibliometri.swepub.kb.se/databearbetning>

¹⁴ *Affiliation & Country* i Appendix 6.1. Kravspecifikation för metadata

¹⁵ *Person identifier & ORCID* i Appendix 6.1. Kravspecifikation för metadata

2.5. Roll och antal upphovspersoner (creator count)

För medräkning och författarfraktionering vid analys behöver ett numeriskt antal upphovspersoner (creator count) anges utifrån rollerna författare ('aut'), redaktörer ('edt') och övriga upphovspersoner ('cre') enligt MARC-relators¹⁶. För att medräkna andra upphovspersoner med andra angivna roller, till exempel fotografer, behöver även MARC-relator 'cre' anges. För korrekt fraktionering är det ytterst viktigt att antal upphovspersoner stämmer oavsett om alla upphovspersoner är namngivna i posten eller inte.

| Datafält | MODS | NR |
|------------------------------------|---|--|
| <i>Creator count</i> ¹⁷ | Mandatory: <i>Creator count</i> (14) | Obligatoriskt för upphovspersoner som skall medräknas i analys: <i>Antal upphovspersoner</i> (2.6) utifrån rollerna 'aut,' 'edt' och 'cre' (se även <i>Roll</i> 2.5) |

2.6. Titel

Titeln anges för att belägga publikationen.

| Datafält | MODS | NR |
|----------------------------|--|--|
| <i>Title</i> ¹⁸ | Mandatory: <i>Title and subtitle</i> (6.) | Obligatoriskt: <i>Titel</i> (3.1) |

2.7. Utgivningsdatum

För att kunna avgöra vilka publikationer som ska ingå i analyser, behöver utgivningsdatum anges för publicerade verk. Utgivningsdatum anges enbart numeriskt som YYYY och för artiklar i dagstidningar som YYYY-MM-DD, alla andra tecken anses vara felaktiga.

| Datafält | MODS | NR |
|---------------------------------------|---|---|
| <i>Publication date and publisher</i> | Mandatory: Publication date and publisher (16) | Se nedan |
| <i>-Publication</i> | Mandatory if publication is formally published: <i>Publication</i> | Obligatoriskt om applicerbart: <i>Utgivningsdatum</i> (3.4) |

¹⁶ MARC Code List for Relators: <http://loc.gov/marc/relators/relaterm.html>

¹⁷ *Creator count* i Appendix 6.1. Kravspecifikation för metadata

¹⁸ *Title* i Appendix 6.1. Kravspecifikation för metadata

| | | |
|---------------------------|--------------------|--|
| <i>date</i> ¹⁹ | <i>date</i> (16.1) | |
|---------------------------|--------------------|--|

2.8. Utgivare

Förlag ligger till grund för beläggning av publiceringskanalen för böcker. Förlag anges som i publikationen tills ett auktoriserat register över publiceringskanaler har upprättats.²⁰

| Datafält | MODS | NR |
|----------------------------------|--|--|
| - <i>Publisher</i> ²¹ | Mandatory if applicable: <i>Publisher</i> (16.2) | Obligatoriskt om applicerbart: <i>Utgivare</i> (3.5) |

2.9. Värdepublikationens titel

Värdepublikationens titel anges för att belägga den publikation där en tidskriftsartikel, konferensoutput eller ett bokkapitel ingår. Titel anges som i publikationen fram tills det att ett auktoriserat register över publiceringskanaler har upprättats.²²

| Datafält | MODS | NR |
|-------------------------------------|---|---|
| <i>Source</i> | Mandatory when the resource is published as a part of another work: <i>Source</i> (17), <i>Journal or book as source</i> (17.1.), <i>Series as source</i> (17.2) | Se nedan |
| - <i>Source title</i> ²³ | Mandatory if applicable: <i>Source title</i> (17.3) | Obligatoriskt om applicerbart: <i>Värdepublikationens titel</i> (3.6) |

¹⁹ *Publication date* i Appendix 6.1. Kravspecifikation för metadata

²⁰ En prototyp till ett auktoriserat register över publiceringskanaler har utarbetats och går att nå via en API: <https://github.com/libris/swepub-sparql/wiki/Auktoriserat-register-%C3%B6ver-publiceringskanaler-och-organisationer>

²¹ *Publisher* i Appendix 6.1. Kravspecifikation för metadata

²² En prototyp till ett auktoriserat register över publiceringskanaler har utarbetats och går att nå via en API: <https://github.com/libris/swepub-sparql/wiki/Auktoriserat-register-%C3%B6ver-publiceringskanaler-och-organisationer>

²³ *Source title* i Appendix 6.1. Kravspecifikation för metadata

2.10. Output

Beroende på syftet med analysen kan vissa outputtyper uteslutas. Vid till exempel en analys av ett visst forskningsämne är det möjligt att endast tidskriftsartiklar och paper i proceedings tas med. För att säkra att publikationer utgör underlag i analyser bör de registreras med specifika outputtyper. Tidigare angivna publikationstyper bör mappas mot de nya outputtyperna.²⁴ Via uttag av dubblettkandidater i tjänsten SwePub för analys och bibliometri kan man jämföra eller upptäcka felaktiga outputtyper.²⁵

| Datafält | MODS | NR |
|-----------------------------|---|---|
| Resource type ²⁶ | Mandatory: <i>Publication type (10.1)</i> , Mandatory: <i>Output type (10.2)</i> | Obligatoriskt: <i>Outputs (3.7)</i> , se även mappning för outputtyper |

2.11. Identifikatorer för output och publiceringskanaler

Identifikatorer för output och publiceringskanaler underlättar identifiering, länkning, åtkomst och dubblettkontroll. ISI ID används dessutom för citeringsanalyser. ISSN skapar kopplingen till auktoritetsregistret över publiceringskanaler. Därmed är det viktigt att ange alla identifikatorer som är tillgängliga för publicerade verk vid registreringstillfället. Observera att en identifikator kan ha blivit tilldelad redan före publicering men att denna kan komma att ändras när materialet är offentliggjort. Identifikatorer bör ligga på rätt nivå, det vill säga identifikatorer för värdpublikationer som ISBN och DOI bör anges för publiceringskanalen och inte för självaste kapitlet/pappret.

| Datafält | MODS | NR |
|-----------------------------------|--|---|
| Resource identifier ²⁷ | Mandatory if the resource has a unique identifier: <i>Resource identifiers (5.)</i> | Obligatoriskt om applicerbart: <i>Identifikatorer för output (3.8) på publikationsnivå (3.8.1)</i> |
| - <i>Source identifier</i> | Mandatory if applicable: <i>Source identifiers (17.4)</i> | Obligatoriskt om applicerbart: <i>Identifikatorer för publiceringskanaler (3.8) på värdpublikationsnivå (3.8.2)</i> |

²⁴ Mappningstabell mellan tidigare publikationstyper och nya outputtyper:

<http://www.kb.se/dokument/SwePub/SwePub-Output-types-codes-and-mapping.pdf>

²⁵ SwePub för analys och bibliometri/databearbetning: <http://bibliometri.swepub.kb.se/databearbetning>

²⁶ *Resource type* i Appendix 6.1. Kravspecifikation för metadata

²⁷ *Resource identifier & Source identifier* i Appendix 6.1. Kravspecifikation för metadata

2.11.1. DOI

- DOI börjar alltid med prefixet 10- .
- Kontrollera om DOI är valid, till exempel med hjälp av shortDOI Service.²⁸

2.11.2. ISI ID

- Artiklar som ingår i ett tidskriftssupplement får nödvändigtvis inte ISI ID.
- Nyligen publicerade tidskriftsartiklar har nödvändigtvis inte hunnit få ISI ID.
- Kontrollera om indexering av tidskriften i Web of Science (WoS) har upphört. Sök fram det senaste år som tidskriften/serien har varit indexerad i WoS.
- Kontrollera om artikeln ingår i serien Lecture Notes in Computer Science som innehåller både tidskriftsartiklar och konferensbidrag. Endast artiklar som har indexerats som tidskriftsartiklar tilldelas ISI ID i detta fall.

2.11.3. ISBN

- Kontrollera om ISBN är valid, till exempel med hjälp av ISBN Converter.²⁹
- På Global Register of Publishers går det att kontrollera korrekta ISBN-komponenter för områdes- och förlagsbeteckning.³⁰

2.11.4. PubMed ID

- Pubmed ID delas även ut till opublicerade verk och är ingen garanti för materialet är sakkunniggranskat.

2.11.5. ISSN

- ISSN ska anges i formatet YYYY-YYYY om publikationen är en tidskriftsartikel.

2.12. Publiceringsstatus

Vid bibliometriska analyser medräknas oftast endast publicerat material oavsett om det finns identifikatorer som DOI eller PubMed ID angivna. Därmed är det viktigt att ange uppdaterad publiceringsstatus och helst ändra i den befintliga posten för att undvika dubletter. Det är frivilligt att ange publiceringsstatus 'Publicerad' om utgivningsdatum finns. Publikationer utan utgivningsdatum måste dock märkas med publiceringsstatus 'Preprint', 'Submitted', 'Accepted', 'In press' eller 'Epub ahead of print/Online first'.

²⁸ ShortDOI Service: <http://shortdoi.org/>

²⁹ ISBN Converter: http://isbn.org/ISBN_converter

³⁰ Global Register of Publishers: <https://grp.isbn-international.org/>

2.13. Åtkomst till fulltext

För att belägga publikationen och för att kunna följa upp publicering i open access-publikationer³¹ ska länk till fulltext anges tillsammans med information om tillgänglighet om den är känd vid granskningstillfället. Öppet tillgängliga forskningspublikationer på internet anges som "free" för att markera att lärosätet inte kontrollerar eller ansvarar för innehållet. Detta gäller till exempel artiklar i open access-tidskrifter (så kallad guld oa).³² En länk till ett repository anses vara beständig för uppföljning. Definitionen "primary" i SwePub MODS indikerar att länken genereras automatiskt från ett lärosätesarkiv när en kopia av en fulltext finns tillgänglig (så kallad grön oa³³).

Om åtkomsten inte är gratis eller den är okänd anges länken utan definition. Vid embargo anges slutdatum i formatet YYYY-MM-DD som betyder att fulltexten är tillgänglig efter det angivna datumet. För närvarande finns det inte möjlighet att ange definition för hybrider.³⁴

Följande verifieringskällor kan användas för att kontrollera om fulltexten är öppet tillgänglig:

- DOAJ³⁵ för oa-tidskrifter
- DOAB³⁶ för oa-böcker
- SHERPA/RoMEO³⁷ för fritt tillgängliga fulltexter
- OpenDOAR³⁸ för öppna arkiv
- ROAD³⁹ med information om oa kopplad till ISSN
- Upphovspersonen med vetskap om länkens tillgänglighet
- QOAM⁴⁰ för att utvärdera kvaliteten på oa-tidskrifter
- Jeffrey Beall's List of publishers och List of standalone journals⁴¹ som är en lista över opålitliga oa-förlag och tidskrifter utarbetad av en amerikansk bibliotekarie

³¹ Med open access (oa) avses vetenskaplig output – tidskriftsartiklar, böcker, konferensbidrag med mera – som görs öppet tillgängliga på internet.

³² Med öppen tillgång i "guld" avses sakkunniggranskad vetenskaplig output som omedelbart publiceras öppet tillgängligt på internet i en oa-kanal.

³³ Med öppen tillgång i "grön" version avses en sakkunniggranskad och redigerad version av vetenskaplig output som redan är publicerad i en traditionell publiceringskanal och som sedan parallellpubliceras i ett digitalt repository, så snart förlaget tillåter det.

³⁴ Hybrid kallas en artikel som publiceras i en låst prenumerationsbaserad tidskrift, men som görs omedelbart öppet tillgänglig.

³⁵ DOAJ – Directory of Open Access Journals: <https://doaj.org/>.

³⁶ DOAB – Directory of Open Access Books: <http://www.doabooks.org/>

³⁷ SHERPA/RoMEO: <http://www.sherpa.ac.uk/romeo/index.php>

³⁸ OpenDOAR – Directory of Open Access Repositories: <http://www.opendoar.org/>

³⁹ ROAD – Directory of Open Access Scholarly Resources (<http://road.issn.org/en>) innehåller bibliografiska poster av tidskrifter, monografiserier, konferensserier och repositoryer.

⁴⁰ QOAM – Quality Open Access Market: <https://www.qoam.eu/>

⁴¹ Scholarly Open Access: <https://scholarlyoa.com/>

| Datafält | MODS | NR |
|--|--|---|
| Location and accessibility ⁴² | Mandatory if applicable: <i>Location and accessibility</i> (18) | Obligatoriskt om applicerbart: <i>Åtkomst till fulltext</i> (3.17) |
| - URL | Mandatory if applicable: <i>URL</i> (18.1) | Se ovan |

2.14. Ämne

För att ingå i ämnesanalyser och underlätta både nationella och internationella jämförelser bör publikationen vara ämnesklassificerad på minst 3-siffernivå och för output inom konst och medicin på 5-siffernivå enligt *Standard för svensk indelning av forskningsämnen* (SCB/UKÄ). Kategorin 'Annan' på 3-siffernivå bör undvikas. Beroende på ämnesklassifikation kan output hamna i olika uttag.

Ämnesklassifikationen ska utgå från innehållet i publikationen men som stöd kan ämnet för värdpublikationen eller upphovspersonens organisationstillhörighet användas. Alternativt kan ämnesklassifikationen för upphovspersonens tidigare verk eller andra publikationsdatabaser och kataloger utnyttjas för jämförelser. Ett exempel på detta är WoS, som ämnesklassificerar både tidskrifter och artiklar.⁴³ I vissa fall kan det vara bra att stämma av ämnesklassifikationen med upphovspersonen eller institutionen alternativt med lärosätesbibliotekens katalogisatorer.

Till hjälp kan även befintliga mappningstabeller användas eller skapas mellan olika ämnesordssystem. Exempel på en befintlig mappningstabell är mappningen mellan Web of Science Subject Headings (WOS SH) och OECD field of science and technology (FOS) classification⁴⁴ som SCB/UKÄ-ämneskategorierna i sin tur är baserade på. Flera lärosäten har börjat analysera ämnen i output utifrån textmining av abstract eller fulltext.⁴⁵

⁴² *Location and accessibility* i Appendix 6.1. Kravspecifikation för metadata

⁴³ Index till Web of Science Subject Areas:

http://incites.isiknowledge.com/common/help/h_field_category_wos.html

⁴⁴ OECD Classification – Web of Science Subject Headings:

http://incites.isiknowledge.com/common/help/h_field_category_oecd_wos.html

⁴⁵ För SCB/UKÄ-ämneskategori utifrån abstract och ISSN, se till exempel Linköpings universitet:

http://www.ep.liu.se/hsv_categories/index.en.asp

| Datafält | MODS | NR |
|--------------------------------------|---|---|
| Subject classification ⁴⁶ | Mandatory: <i>Subject classification according to Swedish research subjects standard (8.1)</i> | Obligatoriskt: <i>Ämne enligt SCB/UKÄ (3.19)</i> |

2.15. Innehållsmärkning

Ett auktoriserat register över publiceringskanaler kommer att avgöra om publikationen är sakkunniggranskad. För hittbarhetens skull kommer dock innehållsmärkningen fortfarande att användas för avgränsning av sakkunniggranskad, övrigt vetenskapligt och/eller populärvetenskapligt.

Till hjälp för att avgöra vilka tidskriftsartiklar som är sakkunniggranskade kan till exempel Thomson Reuters Master Journal List användas.⁴⁷ För seriella resurser kan Ulrichsweb användas.⁴⁸ Om publikationen anges som sakkunniggranskad i publiceringskanalen eller av utgivaren, till exempel artikel i en tidskrift, paper i proceeding eller kapitel i samlingsverk, kan man utgå från den informationen.

⁴⁶ *Subject classification* i Appendix 6.1. Kravspecifikation för metadata

⁴⁷ Master Journal List: <http://ip-science.thomsonreuters.com/mjl/>

⁴⁸ Ulrichsweb: <http://ulrichsweb.serialssolutions.com/>

3. HANTERING AV OFULLSTÄNDIG METADATA

Det finns regler i SwePub som identifierar brister i datakvaliteten och märker upp posterna med olika feltyper. Via tjänsten SwePub för analys och bibliometri går det att söka fram dessa poster för export och lokal åtgärd.⁴⁹

3.1. Utsökning av feltyper

Till stöd för den manuella granskningen kan man göra uttag av poster med ofullständig metadata. Prioritera de åtgärder som har störst påverkan på processen för kvalitetssäkrad metadata. Bearbeta i första hand de poster som har brister i metadata av hög bibliografisk karaktär som är märkta med allvarlighetsgrad 3, vilket exkluderar dem från dubbletthanteringen och från analysunderlag.

- Ange en eller flera organisationer i formuläret för databearbetning för att göra ett uttag av SwePubs feltyper. Lämnas fältet tomt kommer samtliga organisationer med. För att begränsa resultatet kan man ange årtal alternativt lämna fältet tomt för att få med samtliga år. Åren 2012 till 2016 är förinställda då de motsvarar år för tänkbara analyser. Klicka på knappen 'Ofullständig metadata' för att visa en resultatlista.
- För att begränsa resultatet till en viss feltyp kan man välja feltyp i rullgardinsmenyn alternativt klicka på diagrammet 'Distribution av SwePubs feltyper'. I rullgardinsmenyn syns alla feltyper oavsett om dessa förekommer i underliggande data. I diagrammet syns endast förekommande feltyper i underliggande data. Feltyperna är distribuerade utifrån storleksordning och allvarlighetsgrad, som visas med färg. Orange står för feltyper av den högsta allvarlighetsgraden 3, gul för grad 2 och grön för grad 1.
- I resultatlistan laddas de 500 första postraderna utifrån posternas Record ID. Observera att resultatet kan innehålla dubletter om det uppkommer flera bristfälliga datafält utifrån den valda feltypen.

3.2. Export av feltyper

- Lägg till fält till tabellen utifrån behov. Fältet 'Bristfällig data' visar den metadata som systemet har identifierat som ofullständig.
- Klicka på knappen 'Export' för att skicka uttaget som e-post med en länk till fil i valfritt format.

⁴⁹ SwePub för analys och bibliometri/databearbetning: <http://bibliometri.swepub.kb.se/databearbetning>

3.3. SwePub:s feltyper

I tabellen nedan listas feltyperna enligt allvarlighetsgrad i bokstavsordning.

| | |
|--------------------------|---|
| Feltyp | Kort benämning för en viss typ av ofullständig metadata. |
| Beskrivning | Förklaring till feltypen. |
| Benämning | Benämning för feltypen i datamodellen |
| Allvarlighetsgrad | <p>Beroende på hur mycket feltypen påverkar om posten inkluderas i uttag av kvalitetssäkrad data har de delats in i olika grader med siffrorna 1 till 3, där 3 har högst påverkan:</p> <p>Grad 3 – den högsta allvarlighetsgraden Ofullständig metadata av hög bibliografisk karaktär vilket påverkar huruvida posten medtas vid dubbletthantering.</p> <p>Grad 2 – den mellersta allvarlighetsgraden Ofullständig metadata som påverkar vissa typer av bibliometriska analyser men som inte har betydelse för dubbletthanteringen.</p> <p>Grad 1 – den lägsta allvarlighetsgraden Ofullständig metadata som är av lokal karaktär och inte påverkar dubbletthantering.</p> |
| SwePub MODS | Hänvisning till elementen i formatspecifikationen där korrekt sätt att ange data är beskrivet i kodexempel. |

3.3.1. DOI fel

| | |
|-------------|--|
| Feltyp | DOI fel |
| Beskrivning | DOI felskrivet. DOI börjar alltid med prefix 10. |
| Benämning | “DOI format violation” |

| | |
|-------------------|----------------------------------|
| Allvarlighetsgrad | 3 |
| SwePub MODS | Resource identifiers (avsnitt 5) |

3.3.2. ISBN på fel nivå

| | |
|-------------------|--|
| Feltyp | ISBN på fel nivå |
| Beskrivning | ISBN angivet på fel nivå. För outputtyper Kapitel i samlingsverk borde ISBN registreras på värdpublikationsnivå. |
| Benämning | “ISBN at wrong place violation” |
| Allvarlighetsgrad | 3 |
| SwePub MODS | Source identifiers (avsnitt 17.4) |

3.3.3. ISI ID fel⁵⁰

| | |
|-------------------|---|
| Feltyp | ISI ID fel |
| Beskrivning | ISI ID felskrivet. ISI ID skrivs antingen 1) med 15-tecken som börjar med 000, 2) 15-tecken som börjar med A19 och efterföljs med 2 tecken och därefter med 10 siffror. |
| Benämning | “ISI format violation” |
| Allvarlighetsgrad | 3 |
| SwePub MODS | Resource identifiers (avsnitt 5) |

3.3.4. ISSN fel

| | |
|-------------------|---|
| Feltyp | ISSN fel |
| Beskrivning | ISSN felskrivet. ISSN skrivs enbart som numeriskt värde i format YYYY-YYYY. |
| Benämning | “ISSN format violation” |
| Allvarlighetsgrad | 3 |
| SwePub MODS | Resource identifiers (avsnitt 5), Source identifiers (avsnitt 17.4) |

⁵⁰ UT-nummer i Web of Science.

3.3.5. ISSN saknas

| | |
|-------------------|--|
| Feltyp | ISSN saknas |
| Beskrivning | ISSN saknas. Alla output med outputtyp Tidskriftsartikel och angivna som publicerade borde ha ISSN till värdpublikation. |
| Benämning | “Missing ISSN Violation” |
| Allvarlighetsgrad | 3 |
| SwePub MODS | Source identifiers (avsnitt 17.4) |

3.3.6. Konferens – saknad värdpublikation

| | |
|-------------------|---|
| Feltyp | Konferens – saknad värdpublikation |
| Beskrivning | Konferensvärdpublikation saknas. Paper i proceeding bör registreras tillsammans med en värdpublikation. |
| Benämning | “Missing Conference Title Violation” |
| Allvarlighetsgrad | 3 |
| SwePub MODS | Source titel (avsnitt 17.3) |

3.3.7. ORCID fel

| | |
|-------------------|--|
| Feltyp | ORCID fel |
| Beskrivning | ORCID felskrivet. ORCID skall vara skriven i format XXXX-XXXX-XXXX-XXXX. |
| Benämning | “ORCID format violation” |
| Allvarlighetsgrad | 3 |
| SwePub MODS | Person identifier ORCID (avsnitt 13.1.2) |

3.3.8. PersonID fel

| | |
|-------------------|--|
| Feltyp | PersonID fel |
| Beskrivning | PersonID felangivet. Om personID saknas, bör det inte anges med frågetecken (?), streck (-) eller i annat format utan utelämnas. |
| Benämning | “href / local ID violation” |
| Allvarlighetsgrad | 3 |
| SwePub MODS | Person name, identifier and affiliation |

| | |
|--|----------------|
| | (avsnitt 13.1) |
|--|----------------|

3.3.9. PersonID saknas

| | |
|-------------------|---|
| Feltyp | PersonID saknas |
| Beskrivning | PersonID saknas. Identifikator för det egna lärosätets forskare bör anges antingen som lokalt personID eller som ORCID. |
| Benämning | “Missing Local Creator Identifier Violation” |
| Allvarlighetsgrad | 3 |
| SwePub MODS | Person name, identifier and affiliation (avsnitt 13.1) |

3.3.10. Upphov – antal saknas

| | |
|-------------------|--|
| Feltyp | Upphov – antal saknas |
| Beskrivning | Antal upphovspersoner saknas. Antal upphovspersoner (creator count) måste anges. |
| Benämning | “missing creator count” |
| Allvarlighetsgrad | 3 |
| SwePub MODS | Creator count (avsnitt 14) |

3.3.11. Upphov – olika antal

| | |
|-------------------|--|
| Feltyp | Upphov – olika antal |
| Beskrivning | Antal upphovspersoner stämmer inte. Antalet namn på upphovspersoner kan inte vara större än den angivna siffran i antalet upphovspersoner (creator count). |
| Benämning | “creator count mismatch” |
| Allvarlighetsgrad | 3 |
| SwePub MODS | Person name, identifier and affiliation (avsnitt 13.1) & Creator count (avsnitt 14) |

3.3.12. Upphov saknas

| | |
|-------------------|--|
| Feltyp | Upphov saknas |
| Beskrivning | Lokal upphovsperson saknas. Om output är kopplad till det registrerande lärosätet måste det även finnas minst en upphovsperson / organisation kopplad till lärosätet i posten. |
| Benämning | “missing local creator” |
| Allvarlighetsgrad | 3 |
| SwePub MODS | Affiliation (avsnitt 13.1.3) & Record content source (avsnitt 3.1) |

3.3.13. Utgivningsdatum fel

| | |
|-------------------|---|
| Feltyp | Utgivningsdatum fel |
| Beskrivning | Utgivningsdatum felangivet. För offentligt publicerade output skall utgivningsdatum vara skrivet i formatet YYYY eller YYYY-MM-DD för dagstidningsartiklar. |
| Benämning | “Bad DateIssued Violation” |
| Allvarlighetsgrad | 3 |
| SwePub MODS | Publication date (avsnitt 16.1) |

3.3.14. ISBN – fel landkod

| | |
|-------------------|---|
| Feltyp | ISBN – fel landkod |
| Beskrivning | Fel landkod vid ISBN. Den andra komponenten i ISBN-numret har felaktig områdesbeteckning, till exempel gäller 91 för Sverige. |
| Benämning | “ISBN country code Violation” |
| Allvarlighetsgrad | 2 |
| SwePub MODS | Resource identifiers (avsnitt 5) |

3.3.15. ISBN fel

| | |
|-------------------|---|
| Feltyp | ISBN fel |
| Beskrivning | ISBN felskrivet. ISBN bör vara antingen 10-siffrigt eller 13-siffrigt med 978-prefix. |
| Benämning | “ISBN format Violation” |
| Allvarlighetsgrad | 2 |
| SwePub MODS | Resource identifiers (avsnitt 5) |

3.3.16. Publiceringsstatus fel

| | |
|-------------------|--|
| Feltyp | Publiceringsstatus fel |
| Beskrivning | Publiceringsstatus felaktig. Om output med publiceringsstatus ”submitted” har blivit tilldelad identifikatorer, borde publiceringsstatus kontrolleras. |
| Benämning | “Obsolete publication status violation” |
| Allvarlighetsgrad | 2 |
| SwePub MODS | Publication status (avsnitt 12) & Resource identifiers (avsnitt 5) |

3.3.17. Upphov – dubblering

| | |
|-------------------|---|
| Feltyp | Upphov – dubblering |
| Beskrivning | Dubbel angivelse av samma upphovsperson. En upphovsperson bör vara registrerad enbart en gång för samma roll. |
| Benämning | “Duplicate Name Violation” |
| Allvarlighetsgrad | 2 |
| SwePub MODS | Name Part (avsnitt 13.1.1) |

3.3.18. Upphov – olika namnvarianter

| | |
|-------------------|---|
| Feltyp | Upphov – olika namnvarianter |
| Beskrivning | Olika namnvarianter för samma upphovsperson. Om en upphovsperson har flera namnvarianter angivna i olika poster, försvårar det uttag per upphovsnamn. |
| Benämning | “LocalPersonIDMultipleNameVariants” |
| Allvarlighetsgrad | 2 |
| SwePub MODS | Person name, identifier and affiliation (avsnitt 13.1) & Name Part (avsnitt 13.1.1) |

3.3.19. Ämne saknas

| | |
|-------------------|---|
| Feltyp | Ämne saknas |
| Beskrivning | Ämnesklassificering på 3-siffernivå enligt SCB/UKÄ saknas. Poster måste innehålla ämnesklassificering enligt SCB/UKÄ:s standard (tidigare HSV) på minst 3-siffernivå. |
| Benämning | “missing UKÄ/SCB 3-digit subject code” |
| Allvarlighetsgrad | 1 |
| SwePub MODS | Subject classification according to Swedish research subjects standard (avsnitt 8.1) |

4. HANTERING AV DUBBLETTKANDIDATER

Lokala dubletter uppstår när samma publikation har registrerats flera gånger i en enskild organisations publikationsdatabas. Förutom lokala dubletter uppstår även nationella dubletter när olika organisationer levererar metadata för samma vetenskapliga output, så kallade samarbetspublikationer. Skillnaden mellan lokala och nationella dubletter är att lokala dubletter behöver rensas bort medan nationella dubletter ska finnas kvar för att skapa relationer mellan kvalitetssäkrade poster på bibliometrisk nivå.⁵¹

Som dublett i SwePub räknas de poster som beskriver samma output. Posterna delar samma omfång, är författad av samma upphovspersoner och utgiven i samma värdpublikation med samma identifikatorer under samma utgivningstid. Om dubblettkandidater beskriver samma output men där de skiljer sig i åt i någon av de ovan nämnda faktorerna, till exempel en presentation på en konferens som sedan har getts ut som paper i proceeding, räknas dessa som två olika output.

Dubblettkandidaterna bedöms först maskinellt utifrån en algoritm som bygger på en samling regler. Reglerna ger olika poäng eller viktningar. Till exempel ger publiceringskanal och vissa identifikatorer högre poäng än titel och ISBN. När alla regler har körts summeras poängen.⁵² Dubblettkandidater som får en summa större än 1,1 bedöms maskinellt som dubletter. Om den sammanslagna siffran för viktningen är mindre än 0,8 bedöms dubblettkandidater vara enskilda poster. Dubblettkandidater vars sammanslagna viktning ligger mellan spannet 0,8 och 1,1 tillhör en gråzon som behöver granskas manuellt av organisationerna. Felmarginalen är för stor för att systemet ska kunna bedöma dessa dubblettkandidater maskinellt.

SwePubs dubblettantering sker för poster som inte har feltyper av grad 3 – den högsta allvarlighetsgraden.

4.1. Utsökning av dubblettkandidater⁵³

Till stöd för den manuella granskningen kan man göra uttag av dubblettkandidater.

- Gör en utsökning i formuläret för databearbetning. Genom att ange en organisation väljer du både lokala och nationella dubletter för den organisationen.
- För att begränsa uttaget till nationella dubletter väljer du en eller flera specifika organisationer tillsammans med din egen organisation. Genom att lämna fältet tomt väljer

⁵¹ SwePubs datamodell: <http://www.kb.se/libris/SwePub/Datamodell/>

⁵² För en lista på regler som identifierar dubletter, se Swepubs SPARQL-bibliotek: <https://github.com/libris/swepub-sparql/wiki/Dedupliceringsregler>

⁵³ SwePub för analys och bibliometri/databearbetning: <http://bibliometri.swepub.kb.se/databearbetning>

du samtliga organisationer. För att begränsa resultatet kan man ange årtal alternativt lämna fältet tomt för att få med samtliga år. Åren 2012 till 2016 är förinställda då de motsvarar år för tänkbara analyser. Klicka på knappen 'Dubblettkandidater' för att få fram en resultatlista.

- I förhandsvisningen laddas dubblettkandidater parvist upp allteftersom listan skrollas ner. Vid varje par syns organisationen som har levererat posten tillsammans med id. Vid varje par visas även den sammanlagda siffran för viktningen. För att se och jämföra dubblettkandidaterna, klicka på 'Visa'.

4.2. Bedömning av dubblettkandidater

- För att kunna lämna in bedömningar för poster från din organisation behöver du vara inloggad med din SWAMID/SUNET-inloggning. Efter inloggning syns två bedömningsalternativ, 'Samma' eller 'Olika', vid varje rad.⁵⁴ Om posterna är dubletter, klickar du på Samma. Om posterna inte är dubletter, klickar du på Olika. Den manuella bedömningen sparas därefter i SwePub och posterna betraktas som dubletter eller icke-dubletter av systemet beroende på bedömning. Vid nästa omindexering av SwePub kommer posterna att gå vidare i den bibliometriska kvalitetssäkringsprocessen. Redan bedömda dubblettkandidater flyttas längst ner i listan och visar det valda alternativet. Det går att ändra bedömning.
- Observera att eftersom varje post jämförs mot alla poster uppkommer varje par två gånger, en gång för varje lärosäte. De behöver dock endast bedömas en gång.

4.3. Export av dubblettkandidater

- Klicka på knappen 'Export' för att skicka uttaget som e-post med en länk till fil i valfritt format. Eventuella åtgärder av poster i det lokala systemet måste göras separat. Dubblettkandidater vars felaktigheter inte åtgärdas kommer inte med i den bibliometriska kvalitetssäkringsprocessen och ingår inte i uttagen för bibliometriska analyser.

⁵⁴ SWAMID/SUNET-inloggning. <https://www.sunet.se/swamid/>

5. HANTERING AV SAMARBETSPUBLIKATIONER

Publikationsposter som kommer från olika lärosäten men gäller samma vetenskapliga output betraktas som nationella dubletter, så kallade samarbetspublikationer.

Samarbetspublikationer som är bedömda av systemet som nationella dubletter behöver åtgärdas manuellt om det finns motstridigheter mellan dem. Motstridigheter uppstår när vissa nyckelkomponenter som påverkar den bibliometriska datamodellen inte stämmer:

- Fraktioneringen får inte vid sammanslagning av andelar ha ett värde som överstiger 1. Om samma upphovspersoner affilieras till flera lärosäten för samma output stämmer inte fraktioneringen.
- Antal upphovspersoner får inte variera för samma output.
- Utgivningsdatum får inte variera för samma output mer än ett år.
- Publiceringskanaler får inte variera för samma output.
- De identifikatorer för output som finns angivna i båda posterna får inte variera.

5.1. Utsökning av samarbetspublikationer

Till stöd för hanteringen av motstridigheter kan man göra uttag för samarbetspublikationer i tjänsten SwePub för analys och bibliometri. Genom att åtgärda motstridig metadata räknas posterna med i kvalitetssäkringsprocessen och datauttagen av kvalitetssäkrad data.⁵⁵

- Gör en utsökning i formuläret för databearbetning. Börja med att välja din organisation. Då får du alla poster från alla organisationer som har motstridigheter med poster från din organisation. Om du vill koncentrera dig på samarbetspublikationer med vissa organisationer, väljer du din egen organisation samt en eller flera andra. Om du lämnar organisationsfältet tomt, får du samtliga organisationer. För att begränsa resultatet kan man ange årtal alternativt lämna årsfälten tomma för att få med samtliga. Åren 2012 till 2016 är förinställda då de motsvarar år för tänkbara analyser. Klicka på knappen 'Samarbetspublikationer' för att få fram en resultatlista. I förhandsvisningen laddas samarbetspublikationer allteftersom listan skrollas ner.

⁵⁵ SwePub för analys och bibliometri/databearbetning: <http://bibliometri.swepub.kb.se>

5.2. Bedömning av samarbetspublikationer

- För att jämföra och kunna lämna in bedömningar för poster från din organisation behöver du vara inloggad med din SWAMID/SUNET-inloggning.⁵⁶
- Samarbetspublikationerna listas parvist. Vid varje par syns de organisationer som har levererat posten och motsvarande id. Ovanför varje par syns vilka motstridigheter som systemet har identifierat. Samma post kan uppkomma flera gånger om den beskriver samarbetspublikation mellan flera organisationer.
- Observera att det även kan dyka upp par som inte är samarbetspublikationer i detta flöde om dessa maskinellt har identifierats som dubletter, det vill säga om den sammanslagna siffran för viktningen överstiger 1,1. Om den maskinella bedömningen inte stämmer och posterna inte är dubletter/samarbetspublikationer, behöver den överskridas genom att klicka på knappen 'Ej dubletter'. Den manuella bedömningen sparas då i systemet och i fortsättningen kommer posterna betraktas som icke-dubletter/icke-samarbetspublikationer.
- Granska posterna utifrån kriterierna för motstridiga metadata i samarbetspublikationer.⁵⁷ Om du märker att motstridigheterna beror på felaktigheter i posten från din organisation, korrigera posten i din lokala publikationsdatabas. Om du bedömer att motstridigheterna beror på att angivna uppgifter inte stämmer vid den jämförda organisationen, lämnas åtgärderna åt den organisationen.

5.3. Export av samarbetspublikationer

- Klicka på knappen 'Export' för att skicka filen som e-post. Åtgärda motstridiga metadata i den lokala publikationsdatabasen. Vid nästa omindexering av SwePub kommer åtgärdade poster att gå vidare i den bibliometriska kvalitetssäkringsprocessen.

⁵⁶ SWAMID/SUNET-inloggning: <https://www.sunet.se/swamid/>

⁵⁷ Se avsnitt 5.4.

5.4. Motstridigheter

Tabellen nedan visar vilka typer av motstridigheter det finns i metadata om samarbetspublikationer samt hur mycket de påverkar och eventuellt utesluter posterna från att ingå i uttag av kvalitetssäkrad data.

| | |
|-----------------------------|---|
| Typ av motstridighet | Kort benämning för en viss typ av motstridande metadata. |
| Beskrivning | Förklaring till motstridigheten |
| Benämning | Benämning för motstridigheten i datamodellen |
| Allvarlighetsgrad | <p>Beroende på hur mycket motstridigheten påverkar om posten inkluderas i uttag av kvalitetssäkrad data har de delats in i olika grader med siffrorna 1 till 3, där 3 har högst påverkan:</p> <p>Grad 3 – den högsta allvarlighetsgraden Motstridighet av hög prioritet som påverkar huruvida posten medtas vid uttag av kvalitetssäkrad data.</p> <p>Grad 2 – den mellersta allvarlighetsgraden Motstridig metadata i en nyckelkomponent i den bibliometriska datamodellen som kan påverka uttag av kvalitetssäkrad data.</p> <p>Grad 1 – den lägsta allvarlighetsgraden Motstridighet av låg prioritet som endast påverkar vissa typer av uttag av kvalitetssäkrad data.</p> |
| SwePub MODS | Hänvisning till elementen i formatspecifikationen där korrekt sätt att ange data är beskrivet i kodexempel. |

5.4.1. Fraktionering stämmer ej

| | |
|----------------------|--|
| Typ av motstridighet | Fraktionering stämmer ej |
| Beskrivning | Fraktionering stämmer inte. Fraktioneringen är större än 1, det vill säga samma upphovspersoner affilieras till olika organisationer för samma output. |
| Benämning | “AffiliationFractionConflict” |
| Allvarlighetsgrad | 3 |
| SwePub MODS | Affiliation (avsnitt 13.1.3, 13.1.4, 13.2.1) |

5.4.2. Antal upphov olika

| | |
|----------------------|--|
| Typ av motstridighet | Antal upphov olika |
| Beskrivning | Antal upphovspersoner stämmer inte. Creator count visar olika siffror för samma output i olika organisationers poster. |
| Benämning | “CreatorCountConflict” |
| Allvarlighetsgrad | 3 |
| SwePub MODS | Creator count (avsnitt 14) |

5.4.3. DOI olika

| | |
|----------------------|--|
| Typ av motstridighet | DOI olika |
| Beskrivning | DOI stämmer inte. Samma output har registrerats med olika DOI som identifikator. |
| Benämning | “DOIConflict” |
| Allvarlighetsgrad | 2 |
| SwePub MODS | Resource identifiers (avsnitt 5) |

5.4.4. Kanaler olika

| | |
|----------------------|--|
| Typ av motstridighet | Kanaler olika |
| Beskrivning | Publiceringskanalen stämmer inte. Samma output har registrerats med olika publiceringskanaler. |
| Benämning | “PublicationChannelConflict” |
| Allvarlighetsgrad | 2 |
| SwePub MODS | Source (avsnitt 17) |

5.4.5. År olika

| | |
|----------------------|---|
| Typ av motstridighet | År olika |
| Beskrivning | Publiceringsår stämmer inte. Samma output har registrerats med olika utgivningsår som skiljer sig åt mer än ett år. |
| Benämning | “PublicationYearConflict” |
| Allvarlighetsgrad | 2 |
| SwePub MODS | Publication date (avsnitt. 16.1) |

5.4.6. Outputtyper olika

| | |
|----------------------|--|
| Typ av motstridighet | Outputtyper olika |
| Beskrivning | Outputtypen stämmer inte. Samma output har registrerats med olika outputtyper. |
| Benämning | “PublicationTypeConflict” |
| Allvarlighetsgrad | 1 |
| SwePub MODS | Resource type (avsnitt 10) |

5.4.7. Ämnen olika

| | |
|----------------------|---|
| Typ av motstridighet | Ämnen olika |
| Beskrivning | Ämnesklassifikationen skiljer sig avsevärt åt. Samma output har registrerats med olika ämnesklassifikationer som skiljer sig åt på nivå 1 (ensiffernivå). |
| Benämning | “SubjectAmbiguity” |
| Allvarlighetsgrad | 1 |
| SwePub MODS | Subject classification (avsnitt 8.1) |

5.4.8. Titlar olika

| | |
|----------------------|---|
| Typ av motstridighet | Titlar olika |
| Beskrivning | Titel skiljer sig åt. Samma output har registrerats med olika titlar som skiljer sig åt antingen med en liten detalj eller är helt olika. |
| Benämning | “TitleAmbiguity” |
| Allvarlighetsgrad | 1 |
| SwePub MODS | Title and subtitle (avsnitt 6) |

6. APPENDIX

6.1. Kravspecifikation för metadata

För att ingå i uttag för bibliometriska ändamål bör följande krav mötas.

| Datafält | SwePub MODS 2.6 - formatspecifikation | Nationella riktlinjer för beskrivning av vetenskaplig output |
|---|--|---|
| Resource identifier | Mandatory if the resource has a unique identifier: <i>Resource identifiers</i> (5.) | Obligatoriskt om applicerbart: <i>Identifikatorer för output</i> (3.8) på <i>publikationsnivå</i> (3.8.1) |
| Title | Mandatory: <i>Title and subtitle</i> (6.) | Obligatoriskt: <i>Titel</i> (3.1) |
| Resource type | Mandatory: <i>Publication type</i> (10.1), Mandatory: <i>Output type</i> (10.2) | Obligatoriskt: <i>Outputs</i> ⁵⁸ (3.7), se även mappning för outputtyper ⁵⁹ |
| Subject classification | Mandatory: <i>Subject classification according to Swedish research subjects standard</i> (8.1) | Obligatoriskt: <i>Ämne enligt SCB/UKÄ</i> ⁶⁰ (3.19) |
| Person name, identifier and affiliation | Mandatory for originators affiliated with the organisation responsible for entering the record: <i>Names and affiliations</i> (13): All originators related to the organisation registering the record must be represented by a <i>name</i> element each. | Se nedan |
| - <i>Name and affiliation</i> | Mandatory for originator persons related to the organisation responsible for entering the record: <i>Person name, identifier and affiliation</i> (13.1) | Se nedan |
| - <i>Name part</i> | Mandatory: <i>Name Part</i> (13.1.1) | Obligatoriskt: <i>För- och efternamn</i> (2.1): Alla upphovspersoner registreras så långt som möjligt, annars enligt |

⁵⁸ Svensk indelning av vetenskaplig output 2015: http://www.kb.se/dokument/SwePub/Svensk-indelning-av-vetenskaplig-output_2015.pdf

⁵⁹ Mappningstabell mellan tidigare publikationstyper och nya outputtyper:
<http://www.kb.se/dokument/SwePub/SwePub-Output-types-codes-and-mapping.pdf>

⁶⁰ Standard för svensk indelning av forskningsämnen 2011:
http://www.scb.se/sv/_Dokumentation/Klassifikationer-och-standarder/Standard-for-svensk-indelning-av-forskningsamnen-2011/

| | | |
|--------------------------------|---|--|
| | | minimikravet (första, sista och lärosätets egna upphovspersoner) |
| - <i>Affiliation</i> | Mandatory for persons related to the organisation responsible for entering the publication record: <i>Affiliations</i> (13.1.3) | Obligatoriskt för den egna organisationens upphovspersoner. Frivilligt för externa upphovspersoner från Sverige: <i>Affilieringar</i> (2.3) |
| - <i>Organisation</i> | Mandatory when applicable: <i>Organisation as author</i> (13.2) Mandatory for names related to the organisation responsible for entering the publication record: <i>Affiliation</i> (13.2.1) | Obligatoriskt så som beskrivet under Upphovspersoner: <i>Organisation som upphov</i> (2.7) |
| - <i>Country</i> | Mandatory for up to 10 external originators associated with organisations from countries other than Sweden: <i>Affiliation, country</i> (13.1.5) | Obligatoriskt för externa upphovspersoner med affiliering till ett utländskt lärosäte, för upp till 10 personer – för publikationer utanför Web of Science: <i>Affilieringar</i> (2.3), <i>Länder</i> (2.3.4) |
| - <i>Person identifier</i> | Mandatory for originator persons related to the organisation responsible for entering the record: <i>Person name, identifier and affiliation</i> (13.1) | Obligatoriskt för lärosätets egna upphovspersoner: lärosätetslokalt <i>PersonID</i> (2.4.1) |
| - <i>ORCID</i> | Mandatory if applicable. Use for originators with a relation to the organisation entering the record, if the person has an ORCID: <i>Person identifier ORCID</i> (13.1.2) | Obligatoriskt för lärosätets egna upphovspersoner: <i>ORCID</i> levereras i de fall lärosätet använder dessa (2.4.2) |
| Creator count | Mandatory: <i>Creator count</i> (14) | Obligatoriskt för upphovspersoner som skall medräknas i analys: <i>Antal upphovspersoner</i> (2.6) utifrån rollerna 'aut,' 'edt' och 'cre' (se även <i>Roll</i> 2.5) |
| Publication date and publisher | Mandatory: <i>Publication date and publisher</i> (16) | Se nedan |
| - <i>Publication date</i> | Mandatory if publication is formally published: <i>Publication date</i> (16.1) | Obligatoriskt om applicerbart: <i>Utgivningsdatum</i> (3.4) |

| | | |
|----------------------------|---|--|
| - <i>Publisher</i> | Mandatory if applicable: <i>Publisher</i> (16.2) | Obligatoriskt om applicerbart: <i>Utgivare</i> (3.5) |
| - (<i>Place</i>) | Optional: <i>Place</i> (16.3) | Frivilligt: <i>Utgivningsort</i> (3.15) |
| Source | Mandatory when the resource is published as a part of another work: <i>Source</i> (17), <i>Journal or book as source</i> (17.1.), <i>Series as source</i> (17.2) | Se nedan |
| - <i>Source title</i> | Mandatory if applicable: <i>Source title</i> (17.3) | Obligatoriskt om applicerbart: <i>Värdepublikationens titel</i> (3.6) |
| - <i>Source identifier</i> | Mandatory if applicable: <i>Source identifiers</i> (17.4) | Obligatoriskt om applicerbart: <i>Identifikatorer för publiceringskanaler</i> (3.8) på <i>värdepublikationsnivå</i> (3.8.2) |
| Location and accessibility | Mandatory if applicable: <i>Location and accessibility</i> (18) | Obligatoriskt om applicerbart: <i>Åtkomst till fulltext</i> (3.17) |
| - <i>URL</i> | Mandatory if applicable: <i>URL</i> (18.1) | Se ovan |

6.2. Valideringskällor

SwePub använder sig av nedanstående externa källor för att verifiera data inhöstade från lärosäten och andra myndigheter.

| Källa för validering | Beskrivning | Länk |
|---|---|---|
| Countries ontology | För verifiering av landskoder | http://www.w3.org/Consortium/Offices/Presentations/RDFTutorial/rdfs/Countries.owl |
| Danska listan | För mappning av publiceringskanal och ISSN | http://ufm.dk/forskning-og-innovation/statistik-og-analyser/den-bibliometriske-forskningsindikator/autoritetslister |
| DOAB – Directory of Open Access Books | För verifiering av open access-böcker. | http://www.doabooks.org/ |
| DOAJ – Directory of Open Access Journals | För verifiering av open access-tidskrifter. | https://doaj.org/ |
| Finska listan | För mappning av publiceringskanal och ISSN | http://www.tsv.fi/julkaisufoorumi/haku.php?lang=en |

| | | |
|--|--|--|
| ISBN Country codes | För verifiering av ISBN-landskoder | http://everything2.com/title/ISBN+Country+codes |
| LIBRIS | För mappning av publiceringskanal och ISSN | http://libris.kb.se/ |
| Lista över borttagna tidskrifter i DOAJ | För verifiering av open access-tidskrifter före utgallring i DOAJ den 9 maj 2016. | https://doaj.org/ |
| Listor över maskinellt och manuellt bedömda dubblettkandidater i SwePub | För att avgöra vilka publikationsposter som betraktas som dubletter i kvalitetssäkringsprocessen | Regler som identifierar dubletter finns listade i SPARQL-biblioteket: https://github.com/libris/swepub-sparql/tree/master/sparqls/Rules/Deduplication |
| Norska listan | För mappning av publiceringskanal och ISSN | https://dbh.nsd.uib.no/publiseringskanaler/Forside?sok.avansert=true |
| ROAD – Directory of Open Access Scholarly Resources | För verifiering av open access-material. | http://road.issn.org/en |
| shortDOI Service | För verifiering av DOI | http://shortdoi.org/ |
| Språkkod enligt ISO 639-2b/ISO 639-3 | För verifiering av språkkod | http://www-01.sil.org/iso639-3/download.asp |
| Vetenskapsrådets lista över publiceringskanaler i Web of Science | För verifiering av vilka tidskrifter och konferensserier som borde ha ISI ID | Internt dokument |
| Ämneslista utarbetad av Uppsala universitet utifrån Standard för svensk indelning av forskningsämnen 2011 | För verifiering av ämnesklassifikation | http://app.ub.uu.se/epub/categories/index.cfm |

6.3. Tips på källor för maskinell validering

6.3.1. Validering av identifikatorer

| Källa för validering | Beskrivning | Länk |
|----------------------|---|---|
| CrossRef | För beräkning och verifiering av DOI, ORCID m.m. | http://crossref.org/ API: https://github.com/CrossRef/rest-api-doc/blob/master/rest_api.md |
| ID Converter API | För verifiering av PubMed ID | http://www.ncbi.nlm.nih.gov/pmc/tools/id-converter-api/ |
| ISBNdb API | För verifiering av ISBN | http://isbndb.com/api/v2/docs |
| Mendeley API | För verifiering av Scopus ID, DOI, ISBN, ISSN och PubMed ID | http://dev.mendeley.com/overview/how_it_works.html |
| ORCID | För verifiering av ORCID | API: http://members.orcid.org/api Datafil: http://support.orcid.org/knowledgebase/articles/223698-how-do-i-get-the-public-data-file |

6.3.2. Validering av open access

| Källa för validering | Beskrivning | Länk |
|---|---|---|
| DOAB – Directory of Open Access Books | För verifiering av open access-böcker | http://www.doabooks.org/ |
| DOAJ – Directory of Open Access Journals | För verifiering av open access-tidskrifter | https://doaj.org/ |
| OpenDOAR – Directory of Open Access Repositories | För verifiering av öppna arkiv | http://www.opendoar.org/ |
| ROAD – Directory of Open Access Scholarly Resources | För verifiering av open access-tidskrifter och -serier med ISSN | http://road.issn.org/en |
| SHERPA/RoMEO | För verifiering av fritt tillgängliga fulltexter | http://www.sherpa.ac.uk/romeo/index.php |