

Open Access i SwePub 2010 – 2016

1 Introduktion

Som en del av Kungliga bibliotekets (KB) uppdrag att samordna öppen tillgång till svenska forskningspublikationer avser KB att årligen redovisa hur stor andel av de svenska forskningspublikationerna som publiceras med öppen tillgång (Open Access, OA). Redovisningen görs med hjälp av KB:s nationella publikationsdatabas SwePub som löpande samlar in publikationsposter från 42 svenska forskande organisationers lokala publikationsdatabaser. Denna rapport utgör redovisningen för verksamhetsåret 2017 och innefattar statistik för publiceringsåren 2010 – 2016.

Under år 2016 utnyttjade SwePub den nya tjänsten oaDOI för att verifiera publikationernas OA-status, vilket har lett till att OA-andelarna i denna rapport är avsevärt högre än i tidigare redovisningar. Den nya informationen om OA-status som har tillkommit i SwePub leder också till ett större överlapp mellan grön OA och guld respektive hybrid OA, då en artikel som publicerats som guld eller hybrid OA ofta har en sådan licens att den kan återpubliceras i ett repositorium. I denna mätning noterar vi ett genomsnittligt överlapp mellan guld och grön OA på cirka 10 procentenheter, vilket oftast har dragits bort från grön OA i redovisningarna. Se bilagan i avsnitt 9 för definitioner och information om beräkningar.

Då SwePub i dagsläget saknar information om huruvida en publikation är publicerad så att den kräver någon form av betalning eller inloggning för att läsas (sk "toll access") kommer redovisningarna i denna rapport att innefatta en kategori benämnd "Ej OA (eller status okänd)", som alltså kan innefatta både publikationer i prenumerationsmodeller och publikationer där uppgift om OA-status saknas i SwePub.

2 Andel open access i Swepub 2010 – 2016

2.1 Samtliga publikationer

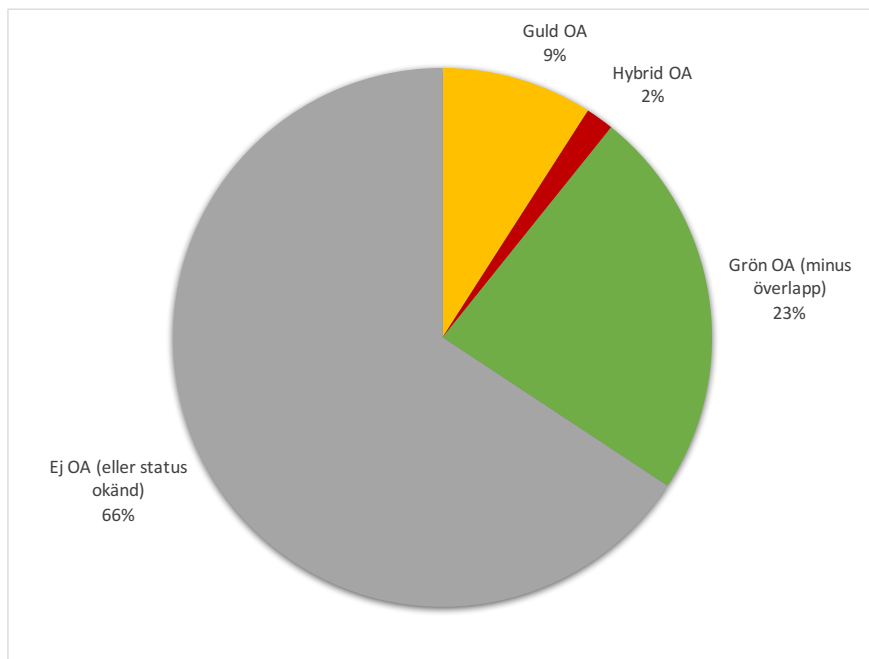


Diagram 1: Andel open access av olika typer för samtliga publikationer i SwePub publiceringsåren 2010 – 2016.

Underlaget för diagram 1 består av samtliga 460 875 publikationsposter i SwePub för publiceringsåren 2010 – 2016 och inkluderar alla dokumenttyper samt både sakkunniggranskat och icke sakkunniggranskat material.

Överlappet av grön OA mot guld och hybrid OA är cirka 7 procentenheter som har räknats bort från andelen grön OA.¹

2.2 Sakkunniggranskade vetenskapliga artiklar och konferensartiklar

Vid studier av utvecklingen av open access brukar man som regel begränsa underlaget till sakkunniggranskade originalartiklar, översiktsartiklar och konferensartiklar. Om man gör den begränsningen i SwePub hittar man 278 195 artikelposter för publiceringsåren 2010 – 2016. Underlagen för diagrammen 2 – 7 är alla baserade på detta urval av publikationer.

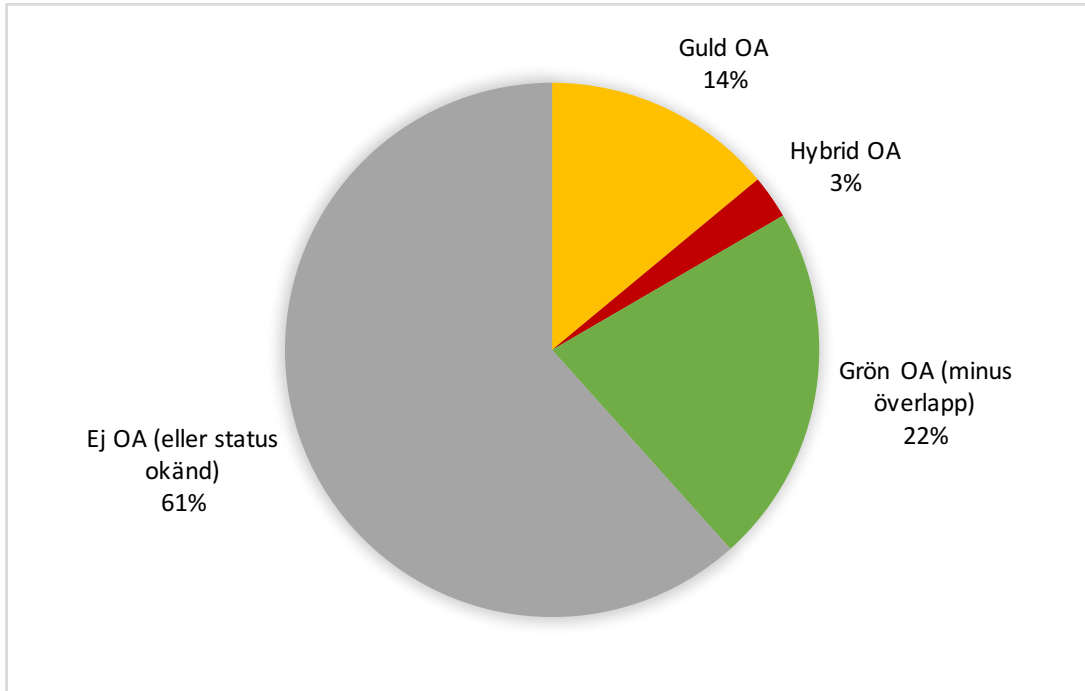


Diagram 2: Andel open access av olika typer för artiklar och konferensartiklar som markerats som sakkunniggranskade i SwePub publiceringsåren 2010 – 2016.

Underlaget för diagram 2 består av 278 195 artiklar och konferensartiklar som markerats som sakkunniggranskade i SwePub. Överlappet av grön OA mot guld och hybrid OA är cirka 11 procentenheter och har här räknats bort från andelen grön OA. Summan OA för de 7 publiceringsåren är cirka 39%.

¹ Publikationer publicerade som guld OA och hybrid OA kan även återpubliceras i repositorer som grön OA och då uppstår en överlappning och dubbelräkning i statistiken, varför överlappet har subtraherats i procentuella sammanställningar. Se bilaga 9.1 för definitioner av de olika kategorierna av OA.

3 Open access-utvecklingen 2010 – 2016

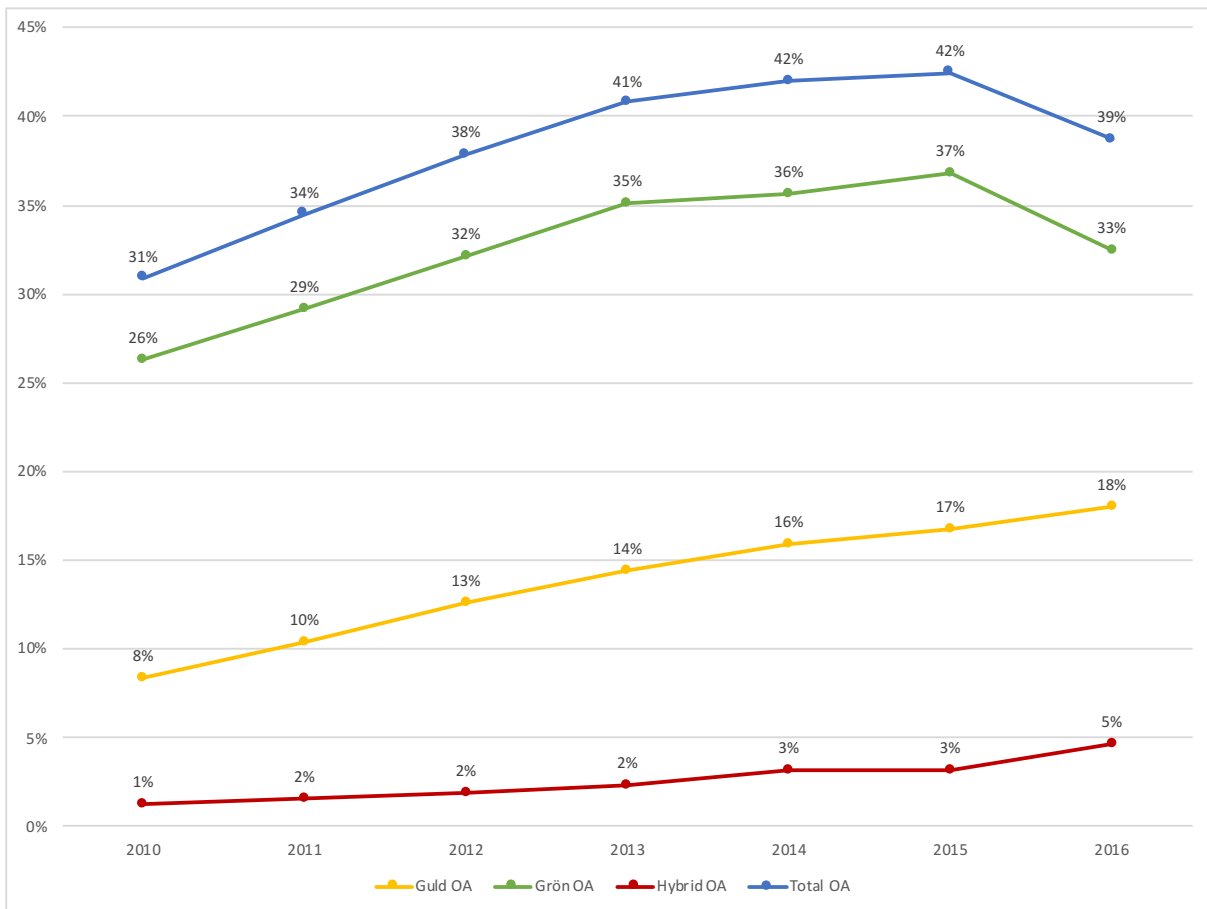


Diagram 3: Trend för andel open access av olika typer för artiklar och konferensartiklar som markerats som sakkunniggranskade i SwePub publiceringsåren 2010 – 2016.

Underlaget för diagram 3 utgörs av 278 195 artiklar och konferensartiklar i SwePub som markerats som sakkunniggranskade.

Överlappet av grön OA mot guld och hybrid OA är i genomsnitt över åren cirka 11 procentenheter, men i diagrammet ovan har inte överlappet subtraherats från grön OA vilket leder till att summan av de olika OA-typerna är större än den totala OA-publiceringen som redovisas i den översta linjen i diagrammet.

Den sjunkande andelen grön OA för publiceringsår 2016 beror troligen på eftersläpningen i registreringen av fulltexter i repositorierna samt embargon för publicering av de parallellpublicerade manuskripten. Den relativt markanta ökningen av hybrid OA publiceringsår 2016 kan bero på det så kallade offset-avtalet Springer Compact som ingicks i mitten av 2016, där svenska forskares artiklar automatiskt publiceras som hybrid OA i Springer Natures tidskrifter och finansieras av universitetsbiblioteken via ett centralt avtal tecknat av Kungliga bibliotekets Bilsamkonsortium.

Siffran 42% total OA i SwePub för publiceringsår 2015 stämmer relativt väl överens med internationella studier, som visar på cirka 45% OA för publiceringsår 2015.¹

4 Forskande organisationers andel open access 2010 – 2016

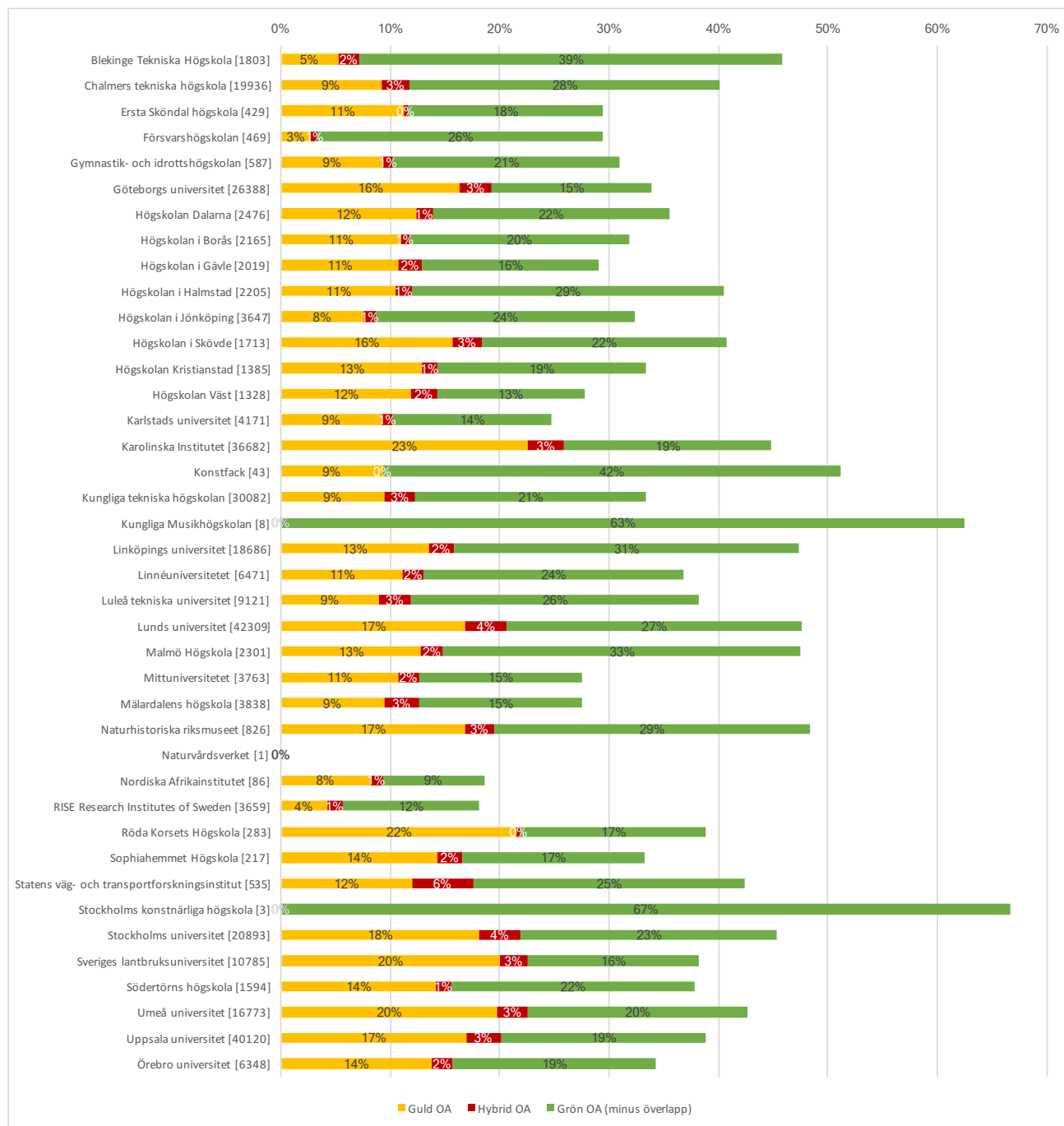


Diagram 4: De SwePub-levererande organisationernas andel open access av olika typer för artiklar och konferensartiklar som markerats som sakkunniggranskade i SwePub publiceringsåret 2010 – 2016. Totalt antal artiklar per organisation anges inom klammer efter organisationsnamnet.

Underlaget för diagram 4 utgörs av 278 195 artiklar och konferensartiklar som markerats som sakkunniggranskade, fördelade över samtliga organisationer som levererar publikationsdata till SwePub.

5 De forskande organisationernas andel guld open access 2010 – 2016

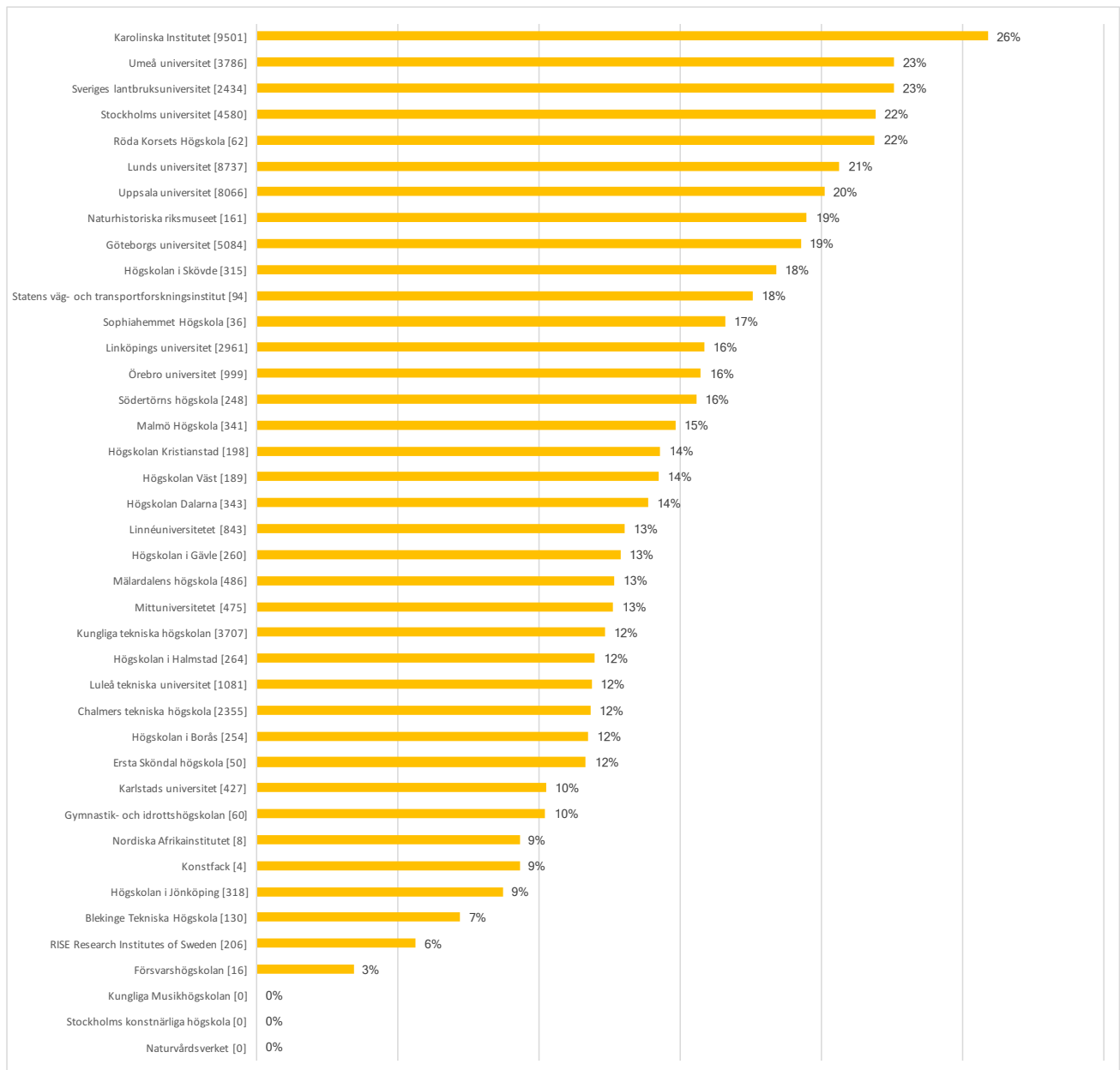


Diagram 5: De SwePub-levererande organisationernas andel guld open access (inklusive hybrid OA) för artiklar och konferensartiklar som markerats som sakkunniggranskade i SwePub publiceringsåret 2010 – 2016. Antal artiklar publicerade som guld/hybrid OA per organisation är angivet inom klammer efter organisationens namn i diagrammet.

Underlaget för diagram 5 utgörs av 278 195 artiklar och konferensartiklar som markerats som sakkunniggranskade, fördelade över samtliga organisationer som levererar publiceringsdata till SwePub.

6 Open access fördelat på forskningsområden 2010 – 2016

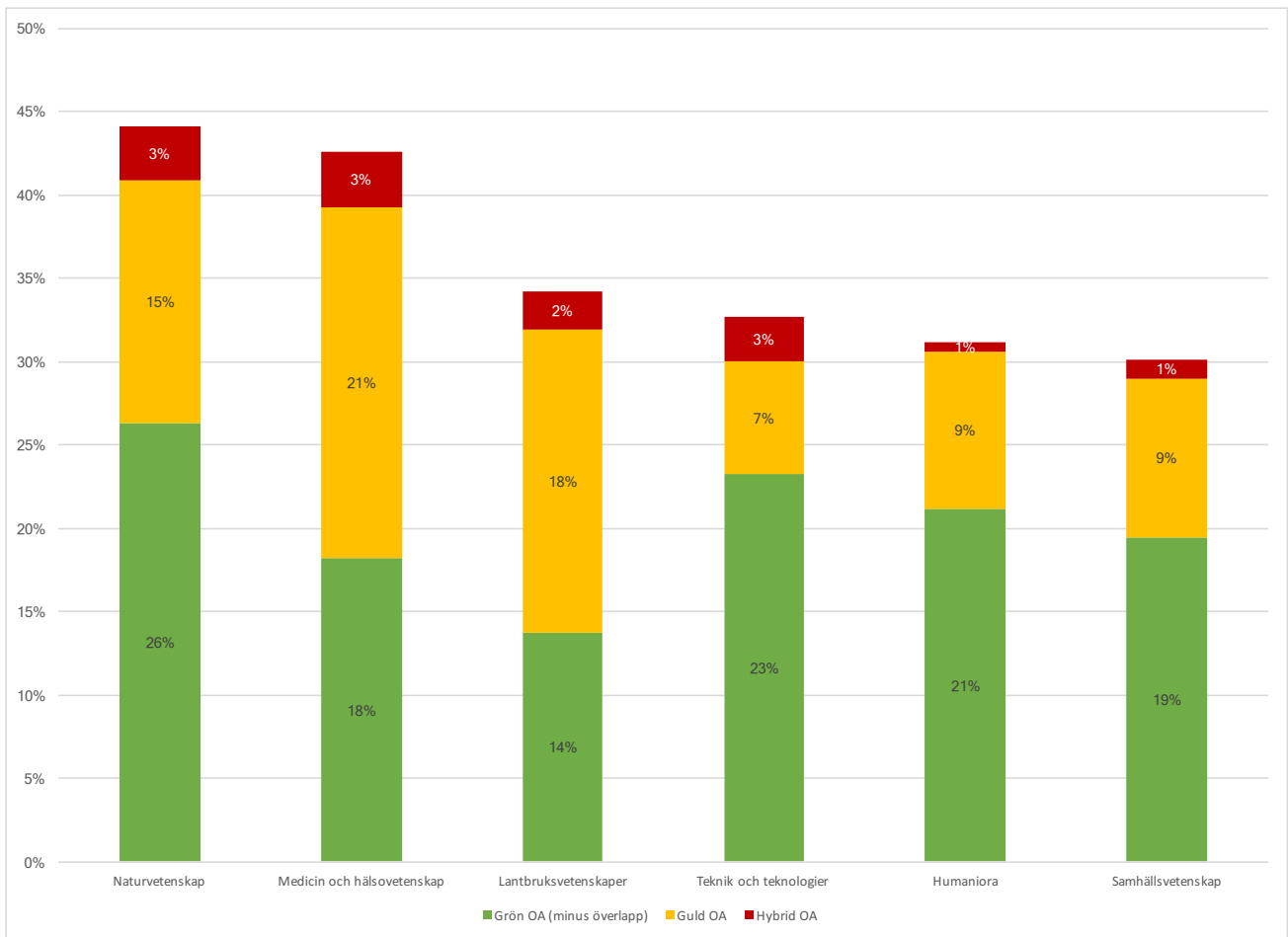


Diagram 6: Open access av olika typer fördelat på forskningsområden (SCB/UKÄ 1-siffernivå) för artiklar och konferensartiklar som markerats som sakkunniggranskade i SwePub 2010 – 2016.

Underlaget för diagram 6 utgörs av 278 195 artiklar och konferensartiklar som markerats som sakkunniggranskade i SwePub.

Vid bedömning av diagram 6 bör man ta i beaktande att cirka 13% av posterna i SwePub saknar information om forskningsområde och att graden av områdesklassning på publikationerna kan variera mellan forskningsområdena. En publikation kan även vara klassificerad inom flera olika forskningsområden, vilket gör att en andel av publikationerna troligen är dubbelräknade i statistiken i diagram 6.

7 Forskningsämnen med högst andel open access 2010 – 2016

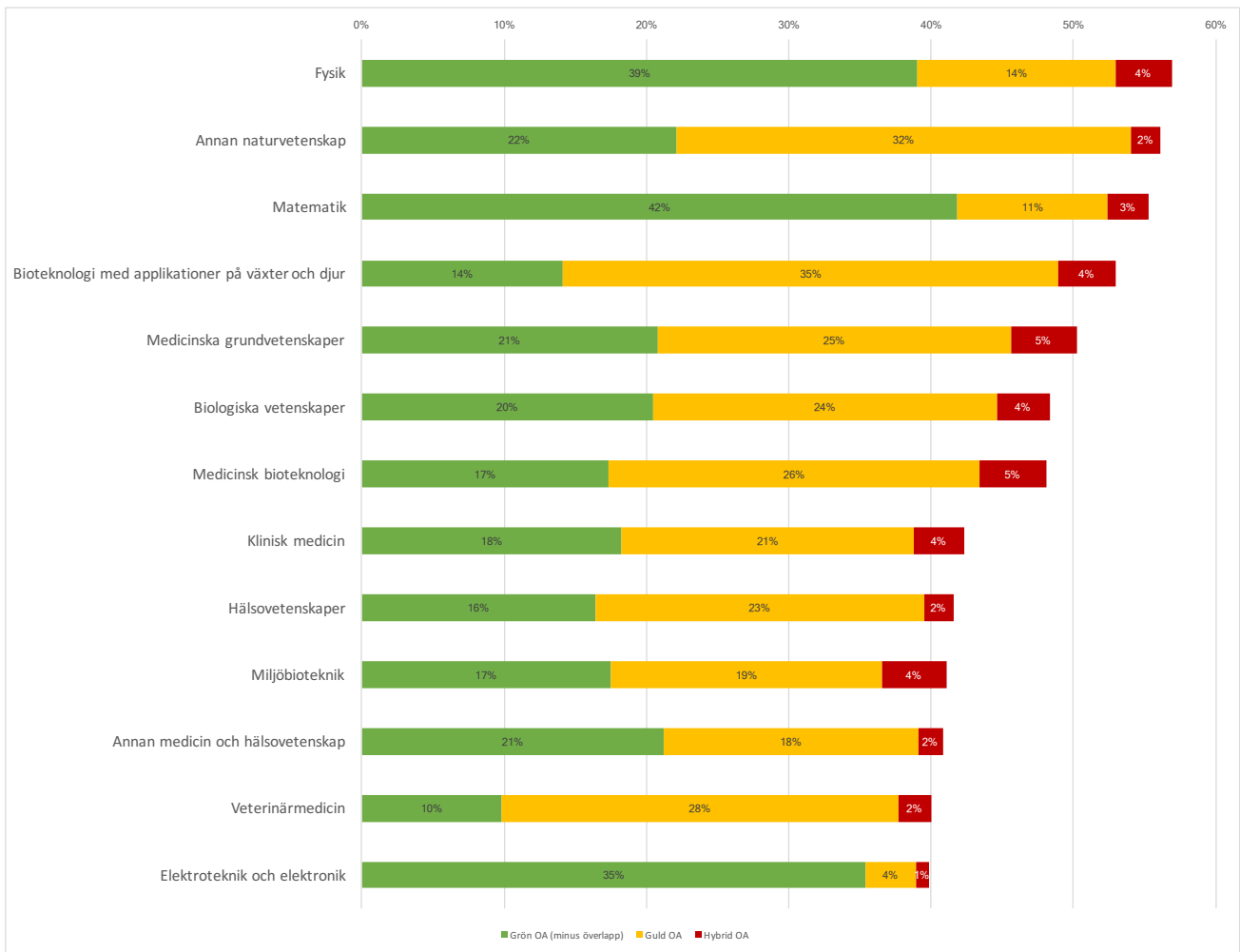


Diagram 7: De 13 forskningsämnen enligt SCB/UKÄ 3-siffrnivå som uppvisar störst andel open access för artiklar och konferensartiklar som markerats som sakkunniggranskade i SwePub 2010 – 2016.

Underlaget för diagram 7 utgörs av 278 195 artiklar och konferensartiklar som markerats som sakkunniggranskade i SwePub.

Vid bedömning av diagram 7 bör man ta i beaktande att cirka 24 procent av posterna i SwePub saknar information om forskningsämne, samt att publikationer kan vara dubbelklassificerade, varvid de kan förekomma i dataunderlaget för flera olika forskningsämnen.

7.1 Open Access fördelat över olika typer av publikationer

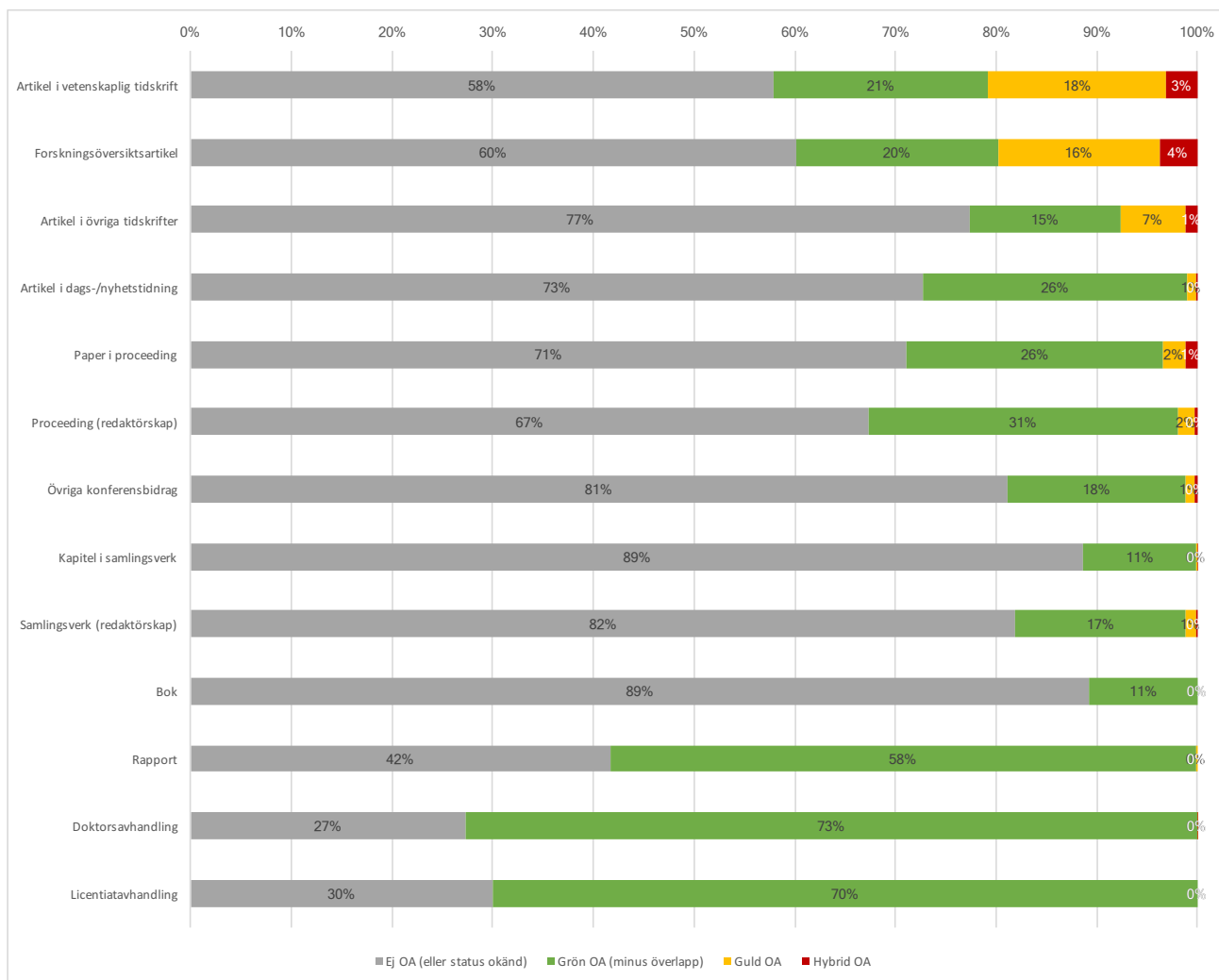


Diagram 8: Andelen open access för 13 olika publikationskategorier i SwePub 2010 – 2016.

Underlaget för diagram 8 är samtliga 460 875 publikationer i SwePub publiceringsåren 2010 – 2016.

8 Referenser

1. Piwowar H, Priem J, Larivière V, Alperin JP, Matthias L, Norlander B, Farley A, West J, Haustein S. (2017). "The State of OA: A large-scale analysis of the prevalence and impact of Open Access articles." *PeerJ Preprints* 5:e3119v1 <https://doi.org/10.7287/peerj.preprints.3119v1>

9 Bilaga: Definitioner, dataunderlag, urval och beräkningar

9.1 Definitioner

Open access (OA): Med open access avses vetenskaplig output – tidskriftsartiklar, böcker, konferensartiklar m.m. – som görs öppet tillgängliga på internet. Om publikationen i någon fas genomgår en sakkunniggranskning är det den sakkunniggranskade versionen av publikationen (ibland i form av ett granskat och godkänt manus) som ska vara fritt tillgänglig. Så kallade pre-prints som publiceras före sakkunniggranskning definieras alltså inte som OA enligt forskningsfinansiärernas krav. Striktare definitioner av OA innefattar även krav på att en publikation ska kunna vidare distribueras och återanvändas (sk. "OA Libre") via en Creative Commons CC-BY-licens, men i denna rapport definieras OA som sk. "OA Gratis", dvs. endast rätten att fritt ladda ned och läsa en publikation är tillräcklig för att den ska räknas som OA.

Guld open access (Guld OA): Med öppen tillgång inom kategorin guld OA avses sakkunniggranskad vetenskaplig output som omedelbart publiceras öppet tillgängligt på internet. I SwePub definieras guld OA som en artikel som är markerad som publicerad som guld OA i metadatat eller är publicerad i en tidskrift som går att verifiera som en ren OA-tidskrift mot DOAJ eller har verifierats som guld OA via tjänsten oaDOI.

Hybrid open access (Hybrid OA) definieras som en artikel som är publicerats öppet tillgänglig på internet, men i en tidskrift som kräver prenumeration eller annan betalning för läsning av övriga artiklar i tidskriften i fråga. Sett på artikelnivå kan hybrid OA definieras som en delmängd av guld OA, då själva artikeln ju är fritt tillgänglig. Begreppet hybrid OA är snarare kopplat till omgivningen artikeln är publicerad i.

Grön open access (Grön OA). Med öppen tillgång som grön OA 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. I SwePub och denna rapport räknas även publikationer som primärpubliceras i ett repository som grön OA. Exempel på sådana publikationer är avhandlingar och rapporter.

Ej OA (eller status okänd): SwePub har för närvarande ingen uppmärkning av sk. "toll access", dvs. artiklar publicerade så att det behövs en prenumeration eller annan betalning eller registrering för att läsa dem. Detta innebär att artiklar som inte är uppmärkta som någon form av OA hamnar i en gemensam kategori benämnd "Ej OA (eller status okänd).

Grön OA överlapp: Eftersom licenserna för artiklar som är publicerade som guld eller hybrid OA tillåter återpublicering i repositoryer kan det hända att dessa artiklar redovisas både som guld/hybrid OA och som grön OA samtidigt, vilket leder till dubbelräkning av dessa artiklar i OA-statistiken. För att undvika dubbelräkning vid sammanställning av statistik har i de flesta sammanställningar överlappet räknats bort från grön OA.

9.2 Dataunderlag

Dataunderlaget för rapporten togs ut från SwePub för Analys² 2017-07-28 och motsvarar totalt 460 875 poster för publikationer åren 2010 – 2016, varav 278 195 dataposter för artiklar, översiktsartiklar och konferensartiklar som markerats som sakkunniggranskade (cirka 60%). Statistiken i denna rapport är som regel baserad på sakkunniggranskade vetenskapliga artiklar och konferensartiklar där inget annat specificeras. Startåret 2010 har valts då det är det år då krav på återrapportering av open access infördes av flertalet svenska forskningsfinansiärer.

Då SwePub är en tjänst som samlar in och aggregerar data från flera olika lokala databaser kan det hända att samma publikation kommer in från flera olika lokala databaser, då forskare från flera olika organisationer har samarbetat kring publikationen och den således har registrerats i flera olika lokala databaser. SwePub utför för närvarande ingen sammanslagning av dessa samarbetsposter. Dessa samarbetspublikationer kommer då att räknas flera gånger i statistiken och leda till en överskattning av antal publikationer i databasen.

I samband med bearbetningen av dataunderlaget för denna rapport har ett visst mått av deduplicering genomförts på så vis att SwePub-poster med en och samma DOI har förts ihop till en post. Denna deduplicering minskade antal poster för publiceringsåren 2010 – 2016 från 514 134 till 460 875, motsvarande cirka 10%. Dock kvarstår en viss dupliceringseffekt för samarbetsposter utan unik identifierare, även om det torde vara en mindre del av det sakkunniggranskade materialet som saknar DOI:er.³

9.3 Dataurval

- Publiceringsår: 2010 – 2016
- Publiceringsstatus: Publicerat material och material utan publiceringsstatus, då vissa lokala system inte förmedlar publiceringsstatus till SwePub.
- Publikationstyper: Konferensartiklar, originalartiklar och översiktsartiklar. I vissa diagram redovisas samtliga publikationstyper.
- Innehållstyp: Sakkunniggranskat material, där annat inte anges.

Den tekniska filtreringen i SwePub:s analysystem Spotfire har gjorts enligt bilderna nedan där inget annat anges.

² <http://bibliometri.swepub.kb.se> och [datadump från ftp://ftp.libris.kb.se/pub/spa/](ftp://ftp.libris.kb.se/pub/spa/)

³ Beräkningar visar att cirka 70% av publikationerna som markerats som sakkunniggranskade i SwePub också har en DOI.

The screenshot shows a search filter interface with the following elements:

- publicationYear:** A dropdown menu showing '2010' and '2016', with '2016' selected.
- Include empty values:** A checkbox that is currently unchecked.
- publicationStatus:** A list of checkboxes with the following options:
 - ACCEPTED
 - EPUB AHEAD OF PRINT/ON...
 - IN PRESS
 - PREPRINT
 - PUBLISHED
 - SUBMITTED
 - (Empty)
- contentType:** A list of checkboxes with the following options:
 - pop
 - ref
 - vet
 - (Empty)
- Other filters (partially visible):**
 - Konferensoutput
 - Paper i proceeding
 - Poster
 - Proceeding (redaktö...
 - Övriga konferensbid...
 - (Empty)
 - Konstnärlig output
 - Publikation
 - Artikel i dags-/nyhet...
 - Artikel i vetenskaplig...
 - Artikel i övriga tidskri...
 - Bidrag till encyklopedi
 - Bok
 - Doktorsavhandling
 - Forskningsöversikts...

Figur 1 Filtrering för urval av publikationer från SwePub som underlag för statistiken för sakkunniggranskade artiklar och konferensartiklar.

9.4 Beräkningar

9.4.1 Kompletterande metadata om OA från oaDOI

Under år 2016 har metadata rörande OA i SwePub kompletterats via uppslag via den nystartade tjänsten oaDOI⁴, som bland annat utnyttjar källorna Directory of Open Access Journals (DOAJ), CrossRef och BASE OA search engine för att verifiera grön, guld och hybrid OA, vilket har lett till att OA-andelarna i denna rapport är avsevärt högre än i tidigare redovisningar. För år 2015 har till exempel andelen grön OA ökat från 17 % i 2016 års mätning till 37 % i årets mätning och andelen guld OA har ökat från 13 % till 17 %.

Den nya informationen om OA-publiceringarna som har adderats till SwePub leder också till ett större överlapp mellan grön OA och guld OA respektive hybrid OA, då en artikel som publicerats som guld OA har en sådan licens att den kan återpubliceras i ett repositorium. I denna mätning noterar vi ett genomsnittligt överlapp på cirka 10 procentenheter.

Då det stora uttaget av metadata rörande OA från oaDOI gjordes under hösten 2016, medan tjänsten fortfarande levererade metadata i version 1 saknas i de flesta fall information både om artiklar som befinner sig bakom betalväggar ("toll access") och artiklar som är publicerade som hybrid OA. Under år 2017 har tjänsten oaDOI vidareutvecklats och levererar nu metadata i version 2 med uppgifter om toll access, hybrid OA och även en ny kategori OA som benämns bronze OA. Bronze OA är artiklar som ligger fria att ladda ned och läsa på förlagens webbplatser, men fortfarande har copyright på förlagen. Under 2017 har oaDOI data av version 1 löpande tillförts SwePub, men bara för de nytilkomna posterna i systemet, varför metadataunderlaget för OA är ojämnt fördelat i databasen. I en studie av OA-status för svenska publikationer i Web of Science som Vetenskapsrådet genomfört under hösten 2017 på färsk data från oaDOI uppvisas en andel hybrid OA straxt över 20% till skillnad mot de 3-4% som visas i denna rapport.

9.4.2 Heltalsräkning kontra fraktionering

Statistiken i denna rapport är baserad på heltalsräkning, så kallade *full counts*, vilket innebär att varje publikation får full vikt vid beräkningar av publikationsantal och medelvärden. Vid bibliometriska studier är det vanligt att man istället använder fraktionsviktning, så kallade *fractional counts*, för att inte publikationer som har författare affilierade till flera analyserade enheter ska räknas flera gånger och få för stor vikt i sammanställningar av medelvärden.

Författarfraktionering innebär att man delar upp en publikation efter antal författare och varje analyserad enhet får en fraktionsvikt motsvarande den andel av det totala antalet författare som hör till enheten. Om man analyser en organisation är detta alltså den andel författare som är affilierade med organisationen i fråga. Om till exempel hälften av författarna är affilierade med en viss organisation får publikationen vikten 0,5 i räkningar av antal och beräkningar av andelar och medelvärden för organisationen i fråga.

Då det verkar föreligga en viss korrelation mellan antal författare och andelen OA finns det risk för att andelen OA i publikationsproduktionen överskattas om man använder full counts. I denna studie har tyvärr inget underlag för fraktionering av publikationerna funnits att tillgå, varför en viss försiktighet vid tolkningen av resultaten är att tillråda. Man kan säga att siffrorna i studien visar hur *stor andel av publikationerna i SwePub* som är OA, men det betyder inte

⁴ <https://oadoi.org/faq>

att de visar hur stor *andel av den svenska artikelproduktionen* som är OA, eftersom man skulle behövt fraktionera publikationerna för att ge en rättvisande bild av detta.

En studie gjord av Vetenskapsrådet på fraktionerade publikationsdata från databasen Web of Science år 2017 visar till exempel att andelen guld OA av den svenska artikelproduktionen åren 2010-2016 är cirka 12%, vilket är lite lägre än de 14% som denna studie visar för andelen guld OA för artiklarna i SwePub motsvarande år. Förutom skillnaden i beräkning baserat på full counts kontra fraktionerade artiklar kan även skillnaden i resultat hänföra sig till att Web of Science täcker ett begränsat urval av sakkunniggranskade svenska artiklar (28 554 år 2016) i jämförelse med SwePub (41 883 år 2016).⁵

⁵ Urvalet i Vetenskapsrådets studie begränsade sig till originalartiklar och översiktsartiklar, och tog alltså inte med konferensartiklar som är inkluderade den övriga statistiken i denna rapport.