

Open Access Indikator for 2021

Del 2

Teknisk beskrivelse af datagrundlag, processer og output Publikationer med publikationsår 2019

0	Forord	2
1	Indledning og Hovedprocesser	3
2	Proces 1: Indsamling af datagrundlag	4
2.1	Universiteternes publikationsdata	4
2.1.1	Krav til universiteter - metadataformat og metode til dataindsamling	4
2.1.2	Årets indgående universiteter og deres forskningsdatabaser	5
2.2	Autoritets- og hjælpedata	5
2.2.1	Directory of Open Access Journals (DOAJ)	5
2.2.2	Sherpa/Romeo (Sh/Ro)	5
2.2.3	Den bibliometriske forskningsindikator (BFI)	5
2.2.4	Autoritetsliste: Accepterede Externe Repositorier ("Whitelist")	6
2.2.5	Autoritetsliste: Tidsskrifter med lange embargoer ("Blacklist")	6
2.3	Årets samlede dataindsamling	6
3	Proces 2: Isolering af publikationer iht. indikatorens genstandsfelt	6
3.1	Genstandsfelt med dubletter	7
3.2	Genstandsfelt uden dubletter	8
4	Proces 3: Beregning af OA realisering og potentiale	9
4.1	Open Access klassifikation - på universitetsniveau	10
4.1.1	Check for Gylden Open Access Potentiale	11
4.1.2	Check for Grøn Open Access Potentiale	12
4.1.3	Check for Uudnyttet og Blokeret Potentiale	15
4.1.4	Check for Open Access Potentiale - Samlet	16
4.2	Open Access klassifikation – på nationalt/hovedforskningsområde niveau	18
5	Proces 4: Kvalitetssikring	18
6	Proces 5: Output	19
6.1	Datarapporter til download	19
6.2	Web-formidling	19
7	Appendix A: Fulltext Download delprocessen	20

Revision 1 fra 6. april 2020

Forord

Danmarks Nationale Strategi for Open Access 2018-2025¹ har som målsætning *”at der fra 2025 og frem er uhindret, digital adgang for alle til alle fagfællebedømte forskningsartikler fra danske forskningsinstitutioner – med maksimalt 12 måneders forsinkelse.”*

Af strategien fremgår endvidere: *”Uddannelses- og Forskningsministeriet vil en gang årligt monitorere implementeringen af Open Access via Open Access Indikatoren (OAI) for at understøtte at alle parter gør deres ypperste for at udvikle og udbrede fri tilgængelighed til danske forskningsresultater.”* Se i øvrigt strategiens tekniske bilag: Baggrund for og monitorering af Open Access Strategi 2018-2025².

Open Access Indikatoren måler på et enkelt publiceringsår inden for strategiens genstandsfelt: *”Videnskabelige artikler og conferencebidrag i tidsskrifter og proceedings med ISSN.”*

Mange tidsskrifter opretholder embargoperioder på 6, 12 eller flere måneder, hvor forskerne er blokeret i at etablere Open Access til deres artikler. Den nationale strategi tillader op til 12 måneders forsinkelse i etableringen af Open Access til en given artikel.

Open Access Indikatoren beregnes én gang årligt og med en forsinkelse som netop sikrer, at eventuelle 12 måneders embargo perioder for publikationer publiceret sidst på året er udløbet på beregningstidspunktet. Således er OAI 2021 beregnet primo marts 2021 på basis af publikationer fra publiceringsåret 2019

Beskrivelsen af Open Access Indikatoren er organiseret i to dele:

- Del 1: Overblik over datagrundlag, processer og output
- Del 2: Teknisk beskrivelse af datagrundlag, processer og output

Henvendelser vedr. indikatoren kan rettes til

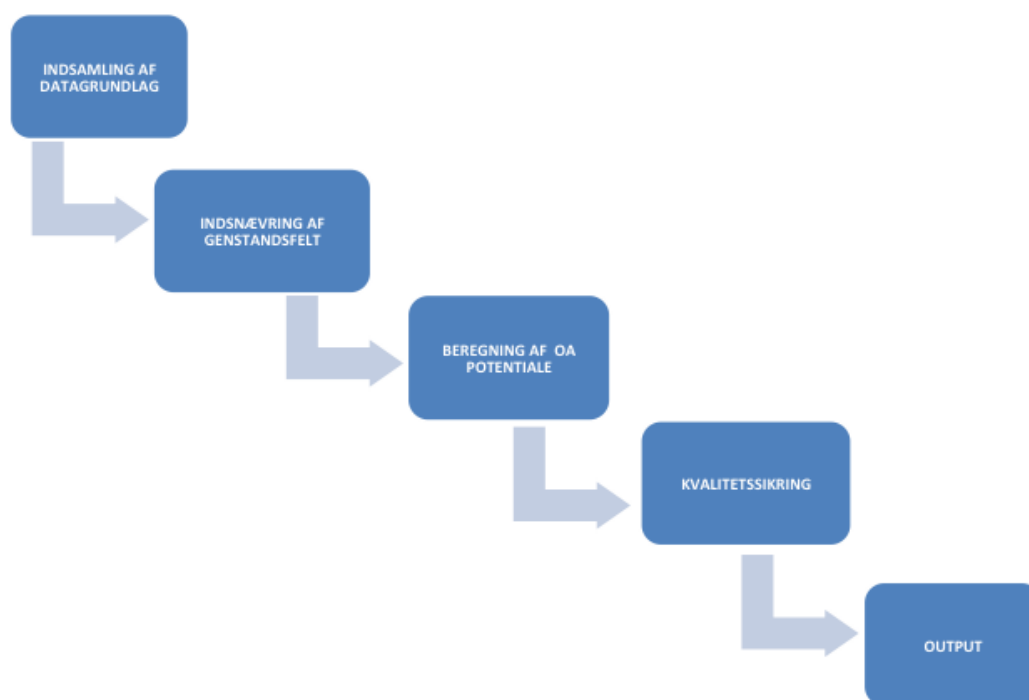
Hanne-Louise Kirkegaard, Chefkonsulent
Uddannelses- og Forskningsstyrelsen
Uddannelses- og Forskningsministeriet
Haraldsgade 53
2100 København Ø
Email: hki@ufm.dk

¹ <https://ufm.dk/forskning-og-innovation/samspil-mellem-viden-og-innovation/open-access/artikler/danmarks-nationale-strategi-for-open-access/final-national-strategi-for-open-access-2018.pdf>

² <https://ufm.dk/forskning-og-innovation/samspil-mellem-viden-og-innovation/open-access/artikler/danmarks-nationale-strategi-for-open-access/teknisk-bilag-til-strategi-baggrund-for-og-monitorering-af-open-access.pdf>

1. Indledning og Hovedprocesser

OA Indikatorens aktiviteter kan struktureres i nedenstående 5 hovedprocesser.

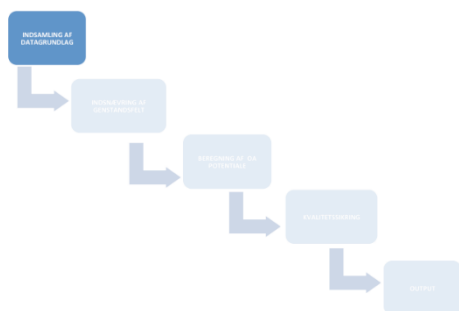


De fem hovedprocesser gennemgås i detaljer i de næste afsnit.

Denne beskrivelse af Open Access Indikatoren henvender sig til et teknisk orienteret publikum og sigter imod at give en dybere indsigt i Indikatorens virkemåde – fra ende til anden samt de enkelte delprocesser.

Beskrivelsen antager, at læseren har basiskendskab til XML samt de basale dele af XPath notation til at referere til XML elementer i et XML dokument der overholder et specifikt XML Schema. Læseren forventes ligeledes at være bekendt med visualiseringer af arbejdsprocesser som workflowdiagrammer.

Proces 1: Indsamling af datagrundlag



Den første aktivitet i OA Indikatoren er indsamling af det samlede datagrundlag, som anvendes i OA Indikatoren og omfatter import af seks nationale og internationale kilder. Datagrundlaget udgøres dels af metadata om universiteternes publikationer og dels af autoritets- og hjælpedata.

2.1 Universiteternes publikationsdata

Metadata om universiteternes publikationer anvendes i Open Access Indikatoren til at etablere genstandsfeltet for indikatoren.

Metadata om universiteternes publikationer indsamles til brug for Open Access Indikatoren én gang årligt. Indsamlingen sker direkte fra universiteterne i et nationalt aftalt XML-baseret metadataudvekslingsformat og med anvendelse af en nationalt aftalt udvekslingsprotokol.

2.1.1. Krav til universiteter - metadataformat og metode til dataindsamling

Et universitet kan indgå i Open Access Indikatoren, såfremt det lever op til følgende minimumskrav:

- Publikationer udgivet af forskere ansat ved universitetet opsamles i universitetets egen forskningsdatabase, som udelukkende rummer information om universitetets egne publikationer, forskere, projekter mv.
- Denne forskningsdatabase skal udstille universitetets publikationsdata vha. OAI-PMH (<http://www.openarchives.org/OAI/openarchivesprotocol.html>) - en standardprotokol som tillader andre at kopiere hele eller dele af databasen.
- Universitetets forskningsdatabase skal understøtte *selective harvesting* og *Sets*, karakteriseret ved deres *setSpec* (kode), som en metode til at kopiere udvalgte dele af databasen.
- Universitetets forskningsdatabase skal udstille et OAI-PMH Set der beskriver al publikations data i databasen og hedder publications:all.
- For dette Set skal universitetets forskningsdatabase understøtte OAI-PMH metadataPrefix "ddf_mxd".
- Indsamles (høstes) data fra dette Set under angivelse af metadataPrefix "ddf-mxd", skal universitetets forskningsdatabase udleverer publikationsdata i DDF-MXD formatet.

2.1.2 Årets indgående universiteter og deres forskningsdatabaser

Følgende 8 universiteter – med tilhørende forskningsdatabaser - indgår i Open Access Indikatoren for 2019 publikationer:

Universitet	Forskningsdatabases OAI-PMH server	Anvendt OAI-PMH setSpec
AAU	https://vbn.aau.dk/ws/oai	publications:all
AU	https://pure.au.dk/ws/oai	publications:all
CBS	https://research-api.cbs.dk/ws/oai	publications:all
DTU	https://backend.orbit.dtu.dk/ws/oai	publications:all
ITU	https://pure.itu.dk/ws/oai	publications:all
KU	https://curis.ku.dk/ws/oai	publications:all
RUC	https://rucforsk.ruc.dk/ws/oai	publications:all
SDU	https://findresearcher.sdu.dk:8443/ws/oai	publications:all

2.2 Autoritets- og hjælpedata

Autoritets- og hjælpedata indsamles til brug for Open Access Indikatoren fra en række kilder. For hver kilde sker indsamlingen én gang årligt. De anvendte indsamlingsmetoder og –formater varierer fra kilde til kilde.

2.2.1 Directory of Open Access Journals (DOAJ)

DOAJ anvendes i Open Access Indikatoren som autoritativ liste over Gylden Open Access tidsskrifter såvel som kilde til data der beskriver tidsskrifternes APC politik.

Parametre for dataindsamlingen:

- Baseret på CSV udtræk af DOAJ data (<https://doaj.org/csv>)

2.2.2 Sherpa/Romeo (Sh/Ro)

Sh/Ro anvendes i Open Access Indikatoren til at bestemme tidsskrifters politik for grøn Open Access og dermed den enkelte tidsskriftartikels Open Access potentiale.

Parametre for dataindsamlingen:

- Anvendt protokol: HTTP (GET fra <http://www.sherpa.ac.uk/downloads/>)
- Dataformat: Properitært XML baseret dataformat (<http://sherpa.ac.uk/news/2012-10-08-RoMEO-API-News.html>)

Note: Sherpa/Romeo har i løbet af 2020 ændret dataformat samt adgang til data i væsentlig grad. Eftersom det nye format har udfaset farvekoderne, som OA Indikatoren har benyttet til at fastslå open access potentialet, er der behov for en fuld analyse af – samt omskrivning af Sherpa/Romeo data. Ifm. 2021 beregningen benyttes derfor Sherpa/Romeo data fra 1. testkørsel juni 2020.

2.2.3 Den bibliometriske forskningsindikator (BFI)

Data fra BFI anvendes i Open Access Indikatoren til to formål

- til at finde dubletter, der skyldes, at samarbejdspublikationer på tværs af universiteter indsamles flere gange, nemlig en gang fra hvert af de samarbejdende universiteter
- til at afgøre eventuelle konflikter mellem sådanne samarbejdspublikations-dubletters angivelse af hovedforskningsområde

Parametre for dataindsamlingen:

- Anvendt protokol: HTTPS (GET fra <https://bfi.fi.dk/AnnualReport>)
- Dataformat: Komprimeret Excel regneark - udokumenteret template

2.2.4 Autoritetsliste: Accepterede Eksterne Repositorier ("Whitelist")

For fuldtekster deponeret i eksterne repositorier anvendes denne autoritetsliste af OA Indikatoren til kun at lade fulltekster deponeret repositorier på listen demonstrere Realiseret Open Access potentiale.

Protokol: Mail (fra autoritetslistens vedligeholdere)

Format: Excel regneark – udokumenteret template

2.2.5 Autoritetsliste: Tidsskrifter med lange embargoer ("Blacklist")

Denne autoritetsliste anvendes af OA Indikatoren til at omklassificere – fra Uudnyttet til Blokeret Open Access potentiale - for publikationer udgivet i tidsskrifter på listen.

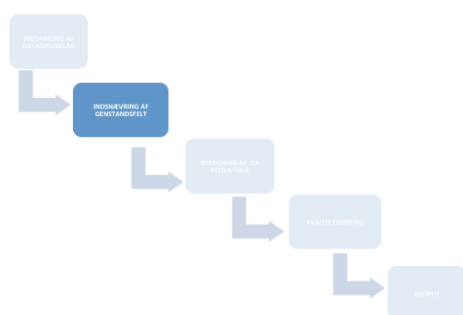
- Protokol: Mail (fra autoritetslistens vedligeholdere)
- Format: Excel regneark – udokumenteret template

2.3 Årets samlede dataindsamling

Dataindsamlingen til Open Access Indikatoren for 2019 kan opsummeres som følger:

Kilde	Protokol	Ver.	Format	Ver.	Indsamlingsdato
AAU	OAI-PMH	2.0	DDF-MXD	1.4.0	3/3 – 2021
AU	OAI-PMH	2.0	DDF-MXD	1.4.0	3/3 – 2021
CBS	OAI-PMH	2.0	DDF-MXD	1.4.0	3/3 – 2021
DTU	OAI-PMH	2.0	DDF-MXD	1.4.0	3/3 – 2021
ITU	OAI-PMH	2.0	DDF-MXD	1.4.0	3/3 – 2021
KU	OAI-PMH	2.0	DDF-MXD	1.4.0	3/3 – 2021
RUC	OAI-PMH	2.0	DDF-MXD	1.4.0	3/3 – 2021
SDU	OAI-PMH	2.0	DDF-MXD	1.4.0	3/3 – 2021
DOAJ	OAI-PMH	2.0	DC+JSON	%	3/3 – 2021
Sh/Ro	HTTP	%	Proprietær	%	9/6 – 2020
BFI	HTTPS	%	Proprietær	%	3/3 – 2021
Whitelist	Mail	%	Proprietær	%	8/6 - 2020
Blacklist	Mail	%	Proprietær	%	19/11 - 2020

3.Proces 2: Isolering af publikationer iht. indikatorens genstandsfelt



Når alle data er indsamlet til OA Indikatoren igangsættes en række aktiviteter, som isolerer de publikationsposter, der tilhører OA Indikatorens genstandsfelt. Dette omfatter ikke alle universiteternes publikationer, men en delmængde heraf, nemlig:

- *Videnskabelige og fagfællebedømte artikler og conferencebidrag i tidsskrifter og proceedings med ISSN*

Følgelig skal denne delmængde isoleres fra den samlede mængde publikationsdata. Dette gøres på to måder for at kunne lave statistik såvel for Danmark som helhed som for det enkelte universitet:

- **Genstandsfelt med dubletter – anvendes til statistik for de enkelte universiteter**
I tilfælde af samarbejdspublikationer på tværs af de indgående universiteter medtages samtlige poster (dubletter) fra de samarbejdende universiteter
- **Genstandsfelt uden dubletter – anvendes til statistikken på nationalt niveau for Danmark som helhed**
I tilfælde af samarbejdspublikationer på tværs af de indgående universiteter medtages kun én post for publikationen

3.1 Genstandsfelt med dubletter

Hver enkelt af kravene til hvornår en publikation tilhører genstandsfeltet, lader sig relativt enkelt oversætte til en regel baseret på DDF-MXD's definition af den tilhørende publikationspost.

Genstandsfeltet er således den mængde af DDF-MXD poster, der overholder kravene ved at opfylde alle reglerne. Reglerne gennemgås nedenfor.

Genstandsfeltet skal kun indeholde publikationsposter med et givent indberetningsår. Den indledende regel er således:

- 0) Publikations **indberetningsår** skal være markeret i posten med en værdi, der svarer til den, som hører til beregningen.
Anvendt regel: Attributten /ddf_doc/@doc_year har værdien (året) hørende til beregningen

Herefter anvendes følgende fem regler på samtlige udleverede publikationsposter:

- 1) Publikationens **type** skal være markeret i posten som "Tidsskriftsartikel", "Review", "Artikel" eller "Konferencebidrag" (samme definition af "artikel" som i BFI).
Anvendt regel: Attributten /ddf_doc/@doc_type har værdien "dja", "djr" eller "dcp"
- 2) Publikationens **review-status** skal være markeret i posten som "Peer-review" (tilsvarende krav gælder for BFI's genstandsfelt).
Anvendt regel: Attributten /ddf_doc/@doc_review har værdien "pr"
- 3) Publikationens **videnskabelige niveau** skal være markeret i posten som "Videnskabeligt" (tilsvarende krav gælder for BFI's genstandsfelt).
Anvendt regel: Attributten /ddf_doc/@doc_level har værdien "sci"
- 4) Publikationens **publiceringsstatus** skal være markeret i posten som "Publiceret" (tilsvarende krav gælder for BFI's genstandsfelt).
Anvendt regel: Attributten /ddf_doc/publication/*/@pub_status har værdien "p"

- 5) Publikationens **publiceringskanal** skal i posten være registreret **med et ISSN**.
Anvendt regel: Elementet /ddf_doc/publication/*/issn har en gyldig ISSN værdi

3.2 Genstandsfelt uden dubletter

For sampublikationer mellem de indgående universiteter kan flere poster i Open Access Indikatoren genstandsfelt med dubletter repræsentere samme publikation. Da dette er uhensigtsmæssigt, når der skal beregnes statistik for hele Danmark, isoleres et genstandsfelt uden dubletter ved at efterbehandle genstandsfeltet med dubletter med en såkaldt dedupliceringsproces. I genstandsfeltet uden dubletter er det således ambitionen, at hver publikation – sampublikation eller ej – kun er repræsenteret med én post.

Open Access Indikatoren deduplicerer ved at danne klynger af dubletposter (poster, der repræsenterer samme publikation), hvor en klynge således repræsenterer én og kun én publikation. Genstandsfelt uden dubletter opstår ved at skabe én post per klynge.

Den anvendte algoritme til at danne klynger er:

- 1) Poster som indgik i BFI beregningen og som i denne blev opfattet som dubletter, føjes til samme klynge
- 2) Poster, hvor signifikante metadata elementer (DOI, titel, undertitel, ISSN, publikationsår, etc.) alle matcher i høj grad, opfattes som repræsenterende samme publikation og føjes til samme klynge

Den anvendte algoritme respekterer BFI's dubletter: Regel (1) sikrer, at poster, som i BFI beregningen blev opfattet som dubletter, også i Open Access indikatoren betragtes som dubletter.

Genstandsfeltet for BFI og for Open Access Indikatoren er imidlertid ikke identiske. Derfor kan andre poster end de, der indgik i BFI beregningen, også optræde som dubletter. Den anvendte algoritme forsøger (best effort) via regel (2) at matche også disse poster sammen til klynger.

Klynger kan således indeholde:

- a. udelukkende poster, der også indgik i BFI beregningen,
- b. både poster der indgik, og poster der ikke indgik i BFI beregningen, eller
- c. udelukkende poster, der ikke indgik i BFI beregningen.

Bemærk for fuldstændighedens skyld, at det under (a) og (b) kan forekomme, at flere dubletklynger fra BFI – efter at være blevet udsat for regel (2) – samles i én og samme klynge i Open Access Indikatoren.

Konfliktløsning

Open Access Indikatoren resultater fordeles på hovedforskningsområder. For at kunne dette skal poster i genstandsfeltet (såvel med som uden dubletter) være markeret med et entydigt hovedforskningsområde.

BFI's definition af hovedforskningsområde anvendes i OA Indikatoren:

- Naturvidenskab/Teknik (sci)
- Samfundsvidenskab (soc)
- Humaniora (hum)
- Medicin (med)

Publikationsposter i genstandsfeltet med dubletter er XML poster i DDF-MXD formatet og indeholder iflg. dette format en entydig markering af hovedforskningsområde. Denne anvendes uden videre.

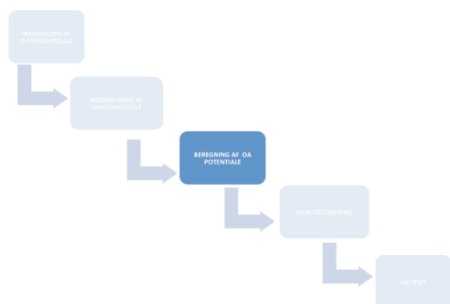
Poster i genstandsfeltet uden dubletter kan stamme fra en klynge af flere poster. Disse poster er ikke nødvendigvis enige om hovedforskningsområdet. I så tilfælde er der tale om en (i BFI-terminologi) hovedforskningsområdekonflikt. Sådanne konflikter skal løses, således at alle klynger (poster) i genstandsfeltet uden dubletter ligeledes har entydigt hovedforskningsområde.

Den anvendte algoritme er:

- 1) Hvis alle klyngens poster har samme hovedforskningsområde, arver klyngen dette hovedforskningsområde.
- 2) Hvis én eller flere af en klynges poster har været en del af en BFI-klynge, tildeles klyngen det samme hovedforskningsområde, som BFI tildelte denne BFI-klynge.
- 3) Hvis ingen af klyngens poster har været en del af BFI beregningen – eller hvis eventuelle BFI poster stammer fra flere BFI-klynger, som ikke er enige om hovedforskningsområde - tildeles klyngen det hovedforskningsområde, som flest af posterne i klyngen er enige om.
- 4) Er der mere end et hovedforskningsområde, som lige mange poster er enige om, tildeles klyngen det af disse hovedforskningsområder, der har størst repræsentation i hele genstandsfeltet.

Open Access Indikatoren løser således sine klyngers eventuelle konflikter vedr. hovedforskningsområde med størst muligt genbrug af BFI's undersøgelser og afgørelser vedr. sådanne konflikter.

4. Proces 3: Beregning af OA realisering og potentiale



Beregningen af OA realisering og potentiale sker ift. grøn og gylden OA og fordelt på universiteter, nationalt og per hovedforskningsområde.

Open Access potentialet – og realiseringen heraf - beregnes publikation for publikation – først på universitetsniveau ud fra genstandsfeltet med dubletter, og dernæst på nationalt og hovedforskningsområde niveau med udgangspunkt i genstandsfeltet uden dubletter.

For begge genstandsfelters vedkommende gælder, at den enkelte publikation klassificeres efter, hvorledes publikationen realiserer Open Access potentialet.

Der anvendes farvekoderne grøn, gul og rød (trafiklys) til at indikere de tre forskellige klassificeringer:

- **Realiseret** Open Access potentiale
- **Uudnyttet** Open Access potentiale, samt
- **Blokeret** Open Access potentiale

For nogle publikationer i genstandsfeltet vil potentialet hvile på muligheden for at downloade en udgave (en fil) som er registreret i metadata posten hørende til publikationen. Af tekniske grunde er download af samtlige sådanne filer én selvstændig delproces, som afvikles før yderligere klassifikation gennemføres. Appendix A indeholder tekniske beskrivelser af denne delproces.

Indikatoren opererer med fire typer af Realiseret Open Access potentiale:

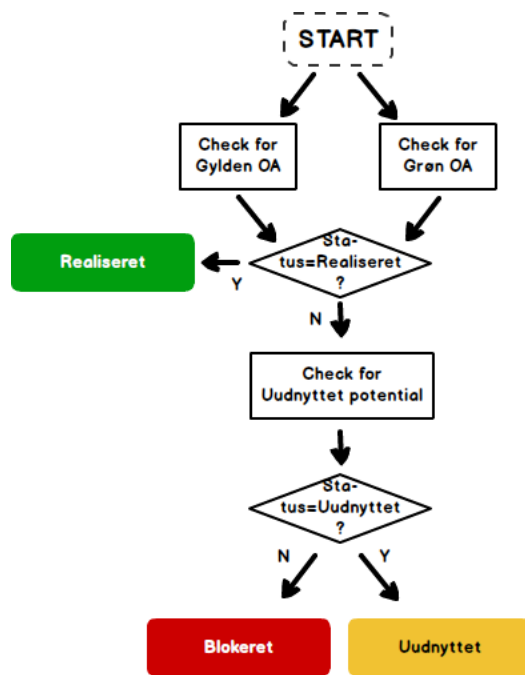
- Gylden Open Access i tidsskrifter med APC
- Gylden Open Access i tidsskrifter uden APC
- Grøn Open Access fra lokalt repository
- Grøn Open Access fra eksternt repository

En publikation der klassificeres med et realiseret Open Access potentiale kan have mere end én type realiseret potentiale.

4.1 Open Access klassifikation - på universitetsniveau

For enhver publikation i genstandsfeltet med dubletter etableres Open Access potentialet – samt dets realisering – ved hjælp af en række checks.

Samlet kan de illustreres ved følgende overblik workflow:



Selvom ovenstående indikerer, at der checkes for Gylden og Grøn OA i parallel, er den praktiske implementering, at Gylden checkes før Grøn.

Hvert af skridtene ovenfor rummer i sig selv workflows. De forklares i større detalje nedenfor.

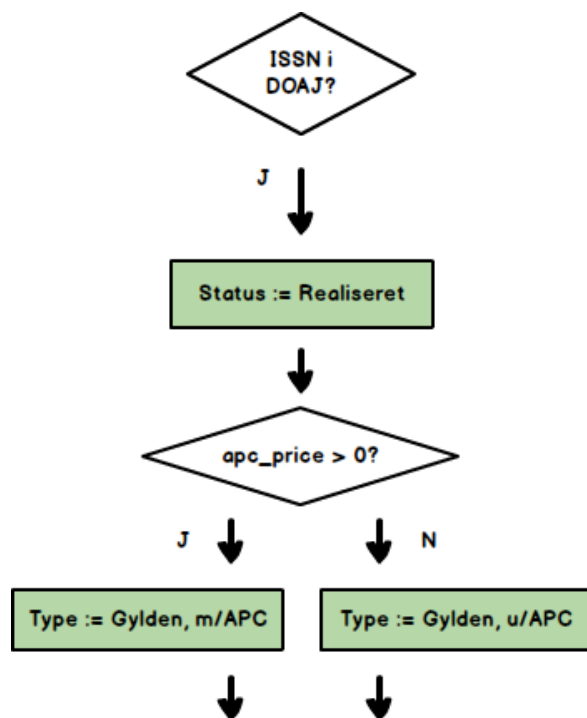
4.1.1 Check for Gylden Open Access potential

Først checkes tidsskriftet (ISSN) imod DOAJ. Findes tidsskriftet her, anses publikationen for at have et (Gyldent) Open Access potentiale, og potentialet anses for Realiseret.

For at bestemme typen af realiseret potentiale, checkes tidsskriftets `apc{average_price}` element. Nedenfor benævnes dette element 'apc_price'.

Hvis `apc_price` har en værdi større end nul, sættes typen af realisering til **Gylden med APC**. Ellers sættes den til **Gylden uden APC**.

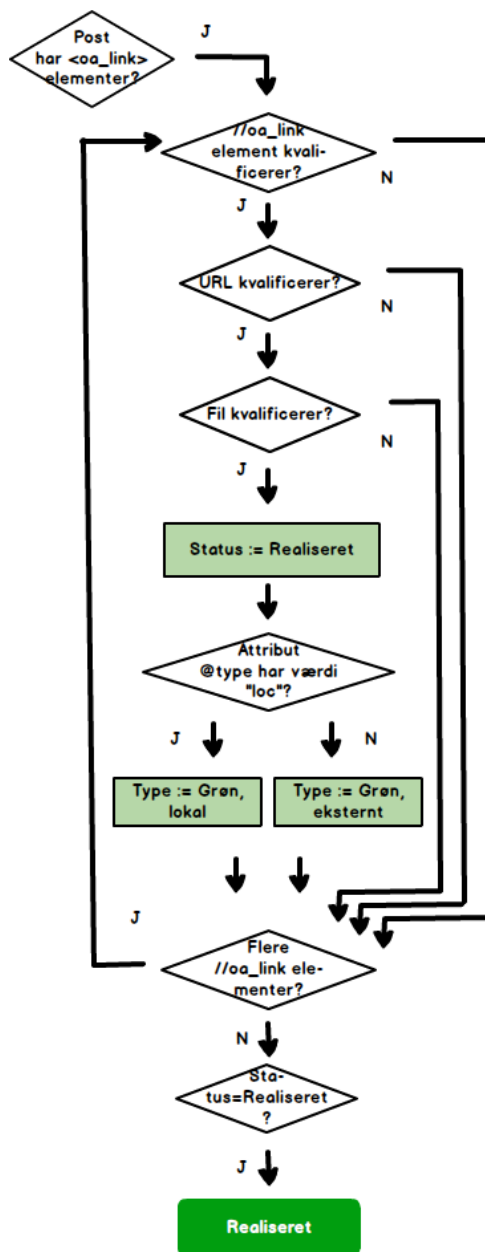
Det tilhørende simple workflow for dette er:



4.1.2 Check for Grøn Open Access potentiale

Grøn Open Access afhænger af metadata elementet /ddf_doc/oa_link i publikationsposten. Nedenfor vil dette element blive refereret med den korte notation //oa_link.

En publikationspost kan indeholde nul, et eller flere //oa_link elementer. Det samlede workflow for at checke for Grøn Open Access er derfor:



Tre beslutningspunkter i dette workflow har at gøre med kvalifikation. Disse tre beslutninger træffes ved hver deres underworkflow, beskrevet nedenfor. For hver fil der med succes passerer alle tre beslutninger, og således klassificerer publikationsposten som Realiseret, bestemmes slutteligt typen af (Grøn Open Access) realisering ved et fjerde beslutningspunkt, som også beskrives nedenfor.

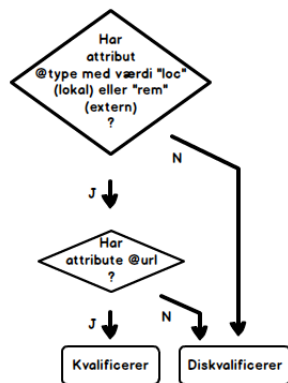
Beslutning 1: //oa_link element kvalificerer?

Et kvalificeret //oa_link element er et //oa_link element hvor

- attribut @type har acceptabel værdi ("loc" for lokal eller "rem" for eksternt – ikke "doi" for DOI)
- attribut @url findes og har værdi

At checke kvalificeringen gøres derfor med følgende lille workflow:

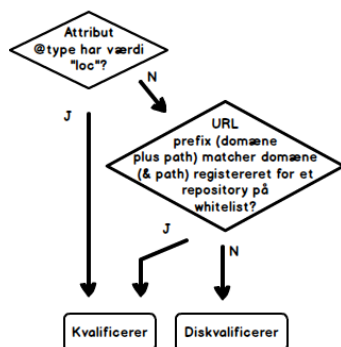
//oa_link element kvalificerer?



Beslutning 2: URL kvalificerer?

En *kvalificeret URL* er enten en URL der peger tilbage til den forskningsdatabase, som metadataposten er modtaget fra (lokal), eller en URL har et prefiks (domænenavn – muligvis sammensat med path) registreret for et repository på autoritetslisten for accepterede eksterne repositorier ("Whitelist"). At checke kvalificeringen gøres derfor med følgende lille workflow:

URL kvalificerer?



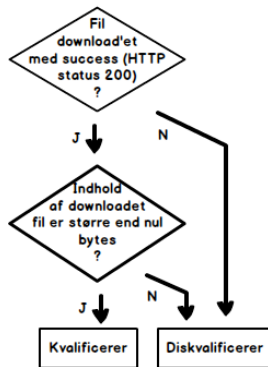
Beslutning 3: Fil kvalificerer?

En *kvalificeret fil* er en fil

- som kan downloades af en computer
- med indhold større end nul

At checke kvalificeringen gøres derfor med følgende lille workflow:

Fil kvalificerer?

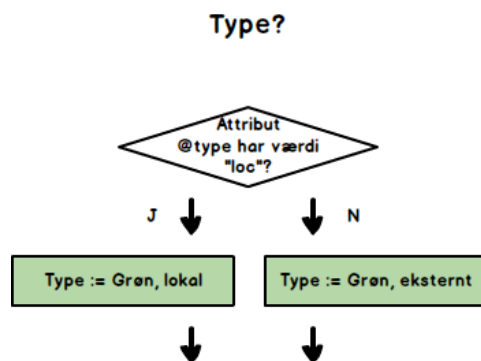


Beslutning 4: Bestemmelse af typen af realisering

Typen af realisering bestemmes af værdien af attributen //oa_link/@type:

- Hvis denne attribute har værdien "loc" sættes typen til **Grøn Open Access fra lokalt repositorie**,
- Ellers sættes den til **Grøn Open Access fra eksternt repositorie**.

Illustreret ved følgende workflow:



4.1.3 Check for Uudnyttet og Blokeret potentiale

Hvis metadataposten ikke har et realiseret Open Access potentiale, undersøges, om potentialet er Uudnyttet. Er det ikke uudnyttet, defineres potentialet som Blokeret.

Til dette anvendes ISSN nummeret registreret i posten, og som publikationen er udgivet i.

Dette anvendes til opslag i Sherpa/Romeo databasen (jvf.

<http://www.sherpa.ac.uk/romeoinfo.html>).

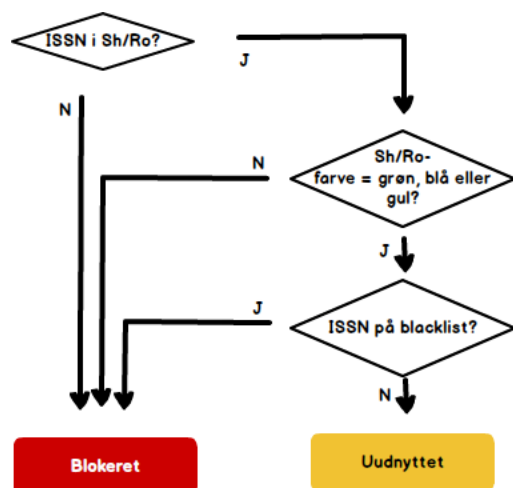
(Se note vedr. Sherpa/Romeo i afsnit 2.2.2.)

Anvendt regel:

- Er tidsskriftets ISSN nummer noteret i Sherpa/Romeo med farvekoderne grøn, blå eller gul, har tidsskriftet OA potentiale, og publikationen anses for at have **Uudnyttet** Open Access potentiale.
 - En undtagelse til denne regel er, hvis tidsskriftets ISSN nummer er noteret for et tidsskrift på Autoritetslisten over Tidsskrifter med Lang Embargo ("Blacklist"). Er det det, anses publikationen for at have et **Blokeret** Open Access potentiale

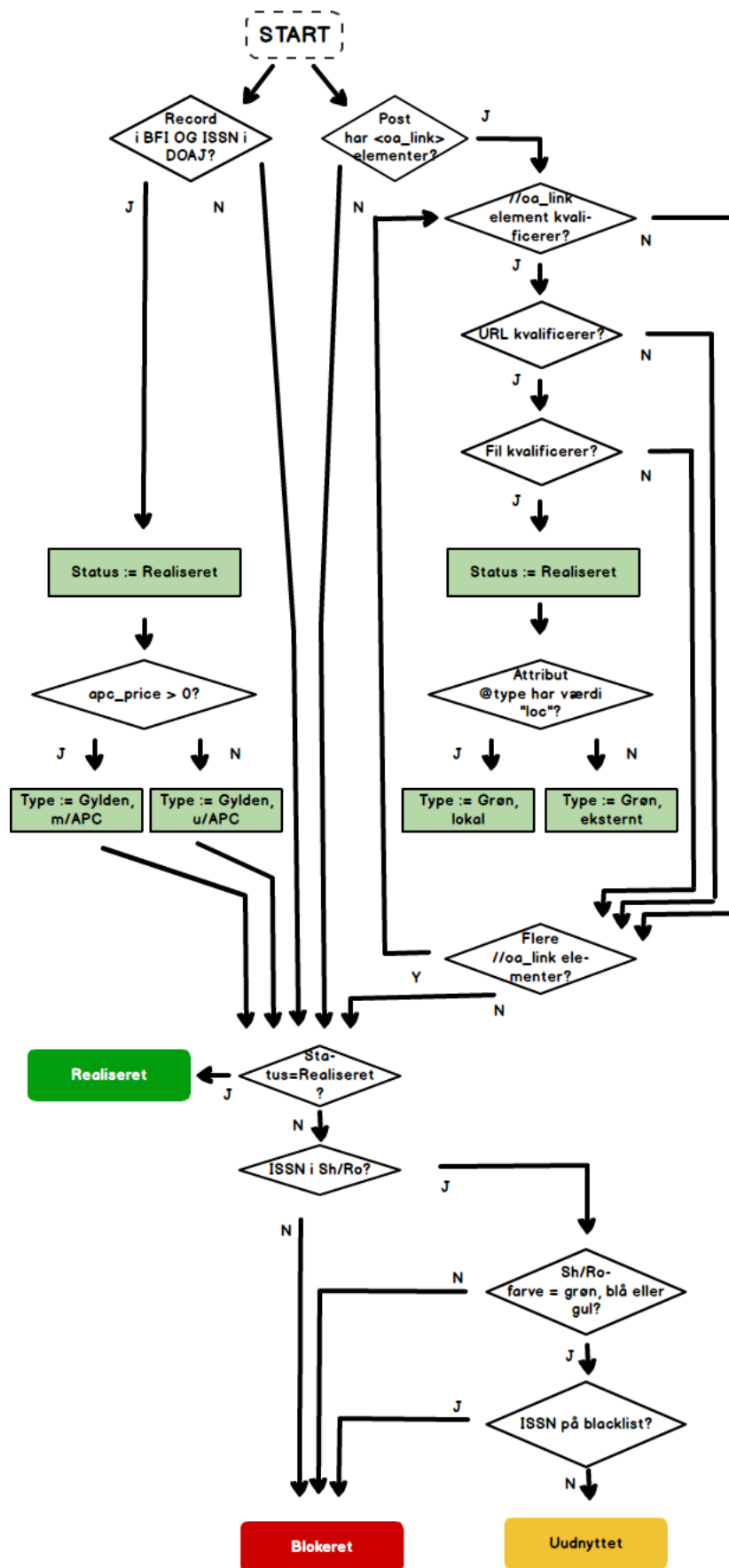
- Er tidsskriftets ISSN nummer ikke noteret i Sherpa/Romeo med farvekoderne grøn, blå eller gul, har tidsskriftet ikke et klart OA potentiale, og publikationen anses for at have **Blokeret** Open Access potentiale.

Dette check svarer til følgende workflow:



4.1.4 Check for Open Access potentiale - Samlet

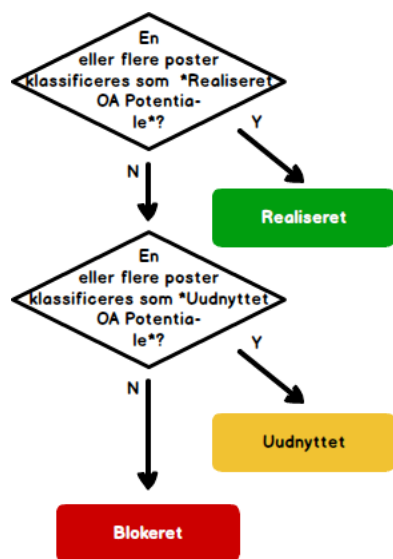
Det samlede beslutningsworkflow for at afgøre Open Access potentialet for en publikationspost er derfor som følger:



4.2 Open Access klassifikation – på nationalt/hovedforskningsområde niveau

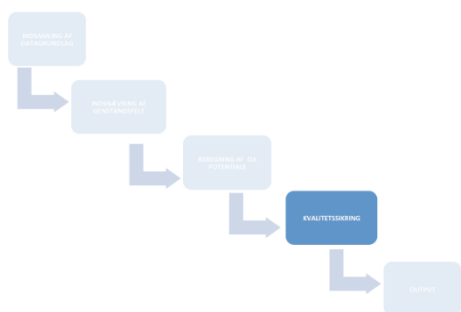
Publikationsposter i genstandsfeltet uden dubletter svarer til klynger hver bestående af én eller flere poster fra genstandsfeltet med dubletter.

Efter at have etableret Open Access potentialet publikation for publikation ved at klassificere publikationerne i genstandsfeltet med dubletter, arver klynger i genstandsfeltet uden dubletter den “**bedst mulige**” klassifikation fra klyngen efter følgende beslutnings workflow:



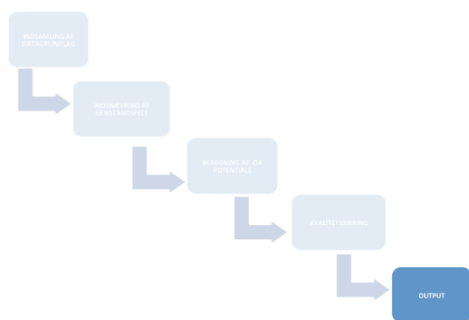
For klynger i genstandsfeltet uden dubletter, der klassificerer som Realiseret, defineres Typen – eller Typerne – af realisering som **foreningsmængden** af Typer fra de af klyngens poster, som klassificerer som Realiseret: Enhver Type realisering knyttet til enhver af klyngens poster med et realiseret potentiale, knyttes også til hele klyngen.

5. Proces 4: Kvalitetssikring



Open Access Indikatorens resultater har været udsat for kvalitetssikring. For en beskrivelse af denne, henvises til overblik dokumentationen.

6. Proces 5: Output



Open Access Indikatoren producerer som output et antal datarapporter samt web-egnede visualiseringer af summeringerne fra disse rapporter.

6.1 Datarapporter til download

Der produceres fem datarapporter:

- 1) Summeringer: Genstandsfelterne optalt samlet og fordelt på **Realiseret**, **Udnyttet** og **Blokeret** Open Access potentiale
 - a. Samlet **Nationalt** (genstandsfelt *uden* dubletter)
 - b. Fordelt på **Hovedforskningsområde** (genstandsfelt *uden* dubletter)
 - c. Fordelt på **de indgående universiteter** (genstandsfelt *med* dubletter)
- 2) Detaljeret grundlag for (a) og (b): Samlet liste over publikationsposter i **genstandsfeltet uden dubletter**
- 3) Detaljeret grundlag for (c): Samlet liste over publikationsposter i **genstandsfeltet med dubletter**
- 4) Den anvendte autoritetsliste over **accepterede eksterne repositorier** ("Whitelist")
- 5) Den anvendte autoritetsliste over **tidsskrifter med lange embargoer** ("Blacklist")

6.2 Web-formidling

Open Access Indikatorens summeringer visualiseres på <https://oaindikator.dk/>, hvor også datarapporter kan downloades.

7. Appendix A: Fuldtekst Download delprocessen

Alle filer (fuldtekster) registreret (med URL) i samtlige poster i genstandsfeltet med dubletter forsøges downloaded i en samlet delproces.

Denne delprocess er implementeret på følgende måde:

- Fuldtekster downloades en ad gangen (serielt; ikke i parallelt)
- Fuldtekster downloades på "server (hostname) Round Robin" måden:
 - en fuldtekst fra server1
 - en fuldtekst fra server 2
 - en fuldtekst fra server 3
 - ...
 - en fuldtekst fra server N
 - en fuldtekst fra server 1
 - en fuldtekst fra server 2
 - ...
 - en fuldtekst fra server N
 - ...
 - ...

Alle downloads foretages automatiseret af OA Indikatorens downloadrobot.

Alle repositorier, som udsættes for disse downloads - lokale (universiteternes forskningsdatabaser) såvel som eksterne – kan identificere netop Indikatorens downloads ved:

- IP Adresse: 130.226.56.102