

Open Access Indikator for 2015

Del 2

Teknisk beskrivelse af datagrundlag, processer og output

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
3.3	Årets genstandsfelter.....	9
4	Proces 3: Beregning af OA realisering og potentiale	10
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	11
4.1.3	Check for Uudnyttet og Uklart Potentiale.....	14
4.1.4	Check for Open Access Potentiale - Samlet.....	15
4.2	Open Access klassifikation – på nationalt/hovedforskningsområde niveau.....	16
5	Proces 4: Kvalitetssikring.....	17
6	Proces 5: Output	18
6.1	Datarapporter til download.....	18
6.2	Web-formidling på Den Danske Forskningsdatabases hjemmeside.....	19
7	Appendix A: Fulltext Download delprocessen.....	20

0 Forord

Den Nationale Styregruppe for Open Access¹ har indstillet til Styrelsen for Forskning og Innovation og Danmarks Elektroniske Fag- og Forskningsbibliotek, at der udvikles en dansk Open Access Indikator. Denne skal støtte implementeringen af den nationale Open Access strategi² - jf. strategiens bemærkninger om monitorering: *"Implementeringen af Open Access skal løbende monitoreres for at sikre, at alle parter gør deres ypperste for at udvikle og udbrede fri tilgængelighed til danske forskningsresultater"*.

Open Access Indikatoren beregnes en gang årligt med genstandsfeltet: *Videnskabelige og fagfællebedømte artikler og konferencebidrag i tidsskrifter og proceedings med ISSN.*

EU kræver i forbindelse med Horizon 2020³, at Open Access realiseres senest 6 måneder efter publicering for hovedområderne naturvidenskab, teknologi og sundhedsvidenskab og senest 12 måneder efter publicering for hovedområderne humaniora og samfundsfag. Denne forsinkelse skyldes, at mange tidsskrifter opretholder såkaldte embargoperioder, hvor de udelukker forskerne fra at etablere Open Access til artiklerne før embargoperiodens udløb.

Da OA Indikatoren beregnes én gang årligt for alle publikationer indenfor genstandsfeltet, er den indrettet til at acceptere et års forsinkelse i Open Access adgangen til publikationerne. Således er OA Indikatoren for 2015 beregnet primo marts 2017 for at tillade et helt års embargoperiode også for publikationer fra december 2014. I praksis betyder dette, at publikationer fra januar 2015 vil kunne have embargoperioder på helt op til 24 måneder og stadig blive godskrevet af OA Indikatoren.

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

Adam Baden/Hanne-Louise Kirkegaard
Styrelsen for Forskning og Uddannelse
Uddannelses- og Forskningsministeriet
Bredgade 40
1260 København K
Email: aba@ufm.dk / hki@ufm.dk

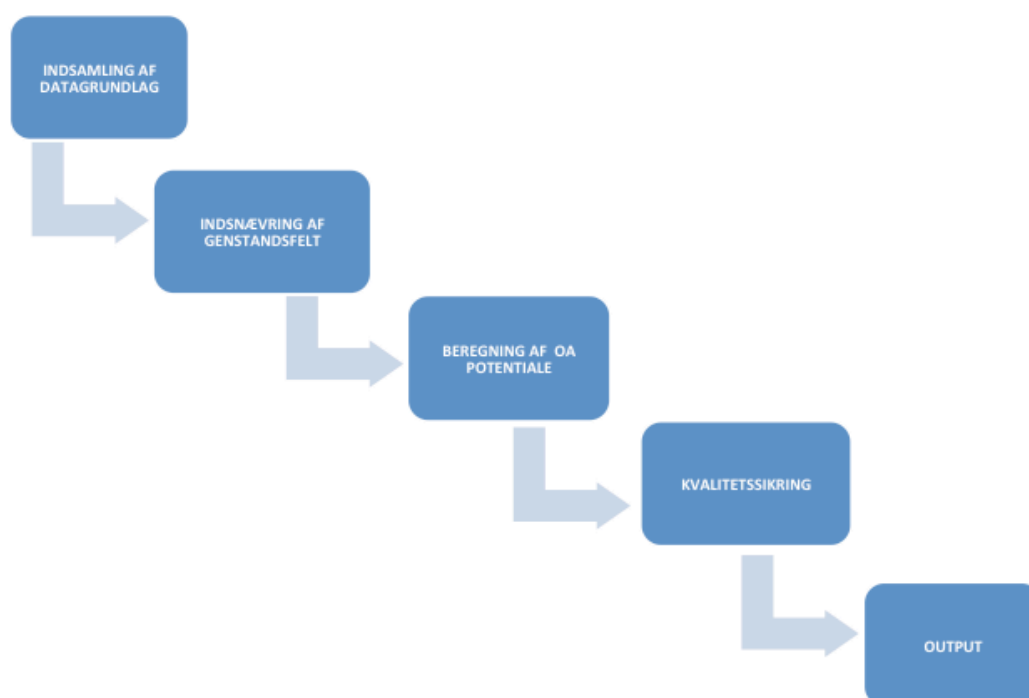
¹ <http://ufm.dk/forskning-og-innovation/samspil-mellem-viden-og-innovation/open-access/artikler/den-nationale-styregruppe>

² <http://ufm.dk/forskning-og-innovation/samspil-mellem-viden-og-innovation/open-access/billeder-og-filer/danmarks-nationale-strategi-for-open-access.pdf>

³ https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.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.

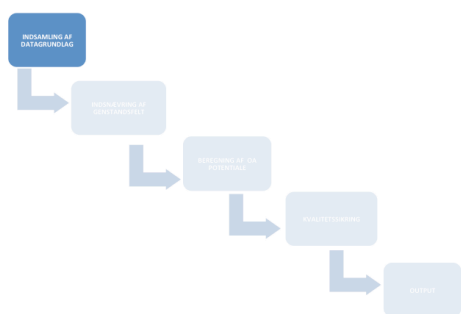
Beskrivelsen antager, at læseren har basiskendskab til XML⁴ samt de basale dele af XPath⁵ notation til at referere til XML elementer i et XML dokument der overholder et specifikt XML Schema. Læseren forventes ligeledes at være bekendt med visualiseringer af arbejdsprocesser som workflow diagrammer⁶.

⁴ <https://www.w3.org/TR/xml/>

⁵ <https://www.w3.org/TR/xpath-30/>

⁶ <https://en.wikipedia.org/wiki/Flowchart>

2 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 alle publikationsdata i databasen
- 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 udlevere publikationsdata i DDF-MXD formatet (<http://mx.forskningsdatabasen.dk/mxd/>).

2.1.2 Årets indgående universiteter og deres forskningsdatabaser

Følgende 8 universiteter – med tilhørende forskningsdatabaser - indgår i Open Access Indikatoren for 2015:

Universitet	Forskningsdatabases OAI-PMH server	Anvendt OAI-PMH setSpec
AAU	http://vbn.aau.dk/ws/oai	publications:all
AU	https://pure.au.dk/ws/oai	publications:all
CBS	http://research.cbs.dk/ws/oai	publications:all
DTU	http://orbit.dtu.dk/ws/oai	publications:all
ITU	https://pure.itu.dk/ws/oai	publications:all
KU	http://curis.ku.dk/ws/oai	publications:all
RUC	http://rucforsk.ruc.dk/ws/oai	publications:all
SDU	http://heinz.sdu.dk:8080/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.

Parametre for dataindsamlingen:

- Anvendt protokol: OAI-PMH (server <http://www.doaj.org/oai/>)
- Anvendt metadataPrefix: oai_dc
- Dataformat: Dublin Core (<http://dublincore.org/documents/dces/>)

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>)

2.2.3 Den bibliometriske forskningsindikator (BFI)

Data fra BFI anvendes i Open Access Indikatoren til tre 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 samarbejdspublications-dubletters angivelse af hovedforskningsområde
- til at sikre, at artikler i DOAJ-validerede tidsskrifter kan anses for at være videnskabelige og fagfællebedømte – dvs. BFI-niveau skal være 1 eller 2.

Parametre for dataindsamlingen:

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

2.2.4 Autoritetsliste: Accepterede Externe Repositorier ("Whitelist")

For fulltekster 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 Uklart Open Access Potential - for publikationer udgivet i tidsskrifter på listen.

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

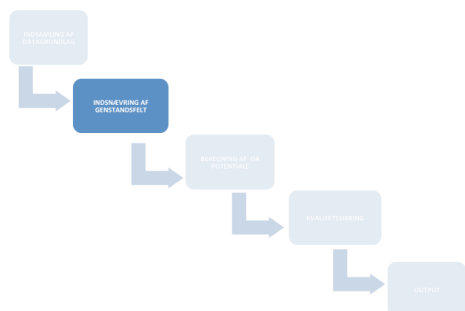
2.3 Årets samlede dataindsamling

Dataindsamlingen til Open Access Indikatoren for 2015 kan opsummeres som følger:

Kilde	Protokol	Ver.	Format	Ver.	Indsamlingsdato	Poster
AAU	OAI-PMH	2.0	DDF-MXD	1.3.0	6/3 – 2017	7248*
AU	OAI-PMH	2.0	DDF-MXD	1.3.0	6/3 – 2017	13221*
CBS	OAI-PMH	2.0	DDF-MXD	1.3.0	6/3 – 2017	2118*
DTU	OAI-PMH	2.0	DDF-MXD	1.3.0	6/3 – 2017	7740*
ITU	OAI-PMH	2.0	DDF-MXD	1.3.0	6/3 – 2017	280*
KU	OAI-PMH	2.0	DDF-MXD	1.3.0	6/3 – 2017	13845*
RUC	OAI-PMH	2.0	DDF-MXD	1.3.0	6/3 – 2017	1550*
SDU	OAI-PMH	2.0	DDF-MXD	1.3.0	6/3 – 2017	7327*
DOAJ	OAI-PMH	2.0	DC	%	6/3 – 2017	13515
Sh/Ro	HTTP	%	Proprietær	%	6/3 – 2017	27032
BFI	HTTPS	%	Proprietær	%	6/3 – 2017	25044
Whitelist	Mail	%	Proprietær	%	26/1 - 2017	15
Blacklist	Mail	%	Proprietær	%	14/12 - 2016	2945

* Med indberetningsår 2015

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 konferencebidrag 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 samarbejdspubliceringer 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 samarbejdspubliceringer 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 fire 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 **publiceringskanal** skal i posten være registreret **med et ISSN**.
Anvendt regel: Elementet /ddf_doc/publication/*/issn har værdi

3.2 Genstandsfelt uden dubletter

For sampublicationer mellem de indgående universiteter kan flere poster i Open Access Indikatoren genstandsfelt med dubletter repræsentere samme publikation. Da dette er u hensigtsmæssigt, 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 – sampublication eller ej – kun er repræsenteret med en post.

Open Access Indikatoren deduplicerer ved at danne klynger af dubletposter (poster, der repræsenterer samme publikation), hvor en klynge således repræsenterer en og kun en publikation. Genstandsfelt uden dubletter opstår ved at skabe en 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 dublet klynger fra BFI – efter at være blevet udsat for regel (2) – samles i en 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.

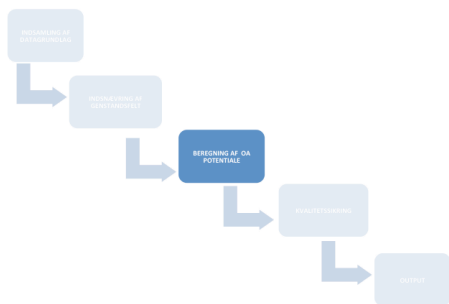
Open Access Indikatoren løser således sine klyngers eventuelle konflikter vedr. hovedforskningsområde med størst muligt genbrug af BFI's undersøgelser og afgørelser vedr. sådanne konflikter.

3.3 Årets genstandsfelter

Datasæt	Poster
Samlet antal publikationsposter indsamlet fra de indgående universiteter	53.429
Genstandsfelt med dubletter	25.070
Genstandsfelt uden dubletter	22.666

For yderligere detaljer, se afsnit om datarapporter.

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:

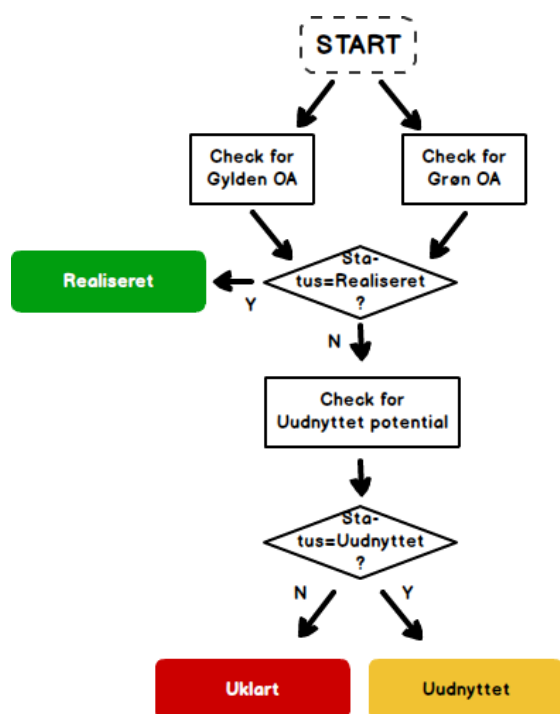
- **Realiseret** Open Access potentiale
- **Uudnyttet** Open Access potentiale, samt
- **Uklart** Open Access potentiale

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 een selvstændig delproces, som afvikles før yderligere klassifikation gennemføres. Appendix A indeholder tekniske beskrivelser af denne delproces.

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:



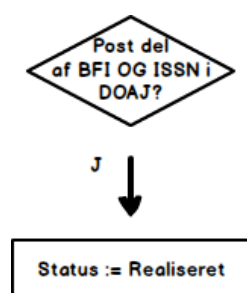
Selvom ovenstående indikerer, at der checkes for Gylden og Grøn OA i parallel, er den praktiske implementation, at Gylden checkes før Grøn.

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

4.1.1 Check for Gylden Open Access Potentiale

Først checkes tidsskriftet imod DOAJ. Findes tidsskriftet her, og har publikationsposten opnået niveau 1 eller 2 i BFI beregningen, anses publikationen for at have et (Gyldent) Open Access potentiale, og potentialet anses for Realiseret.

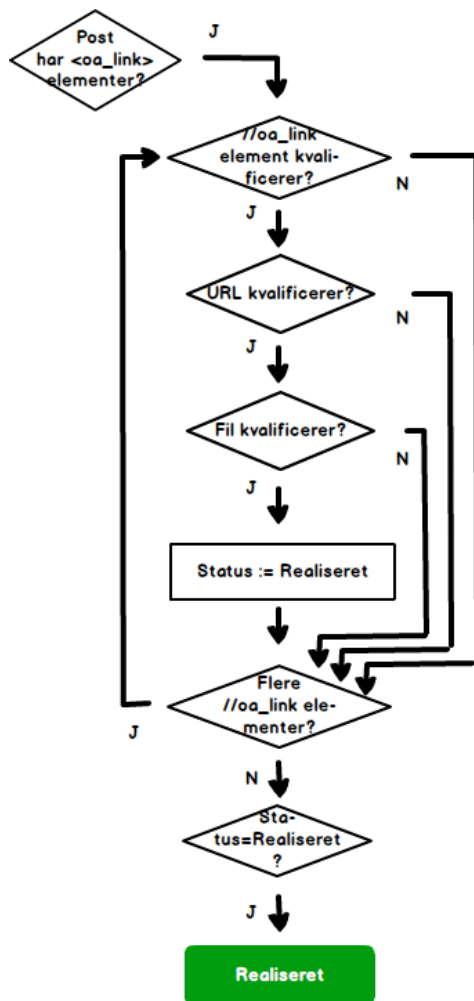
Det tilhørende simple workflow for dette er:



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 under workflow:

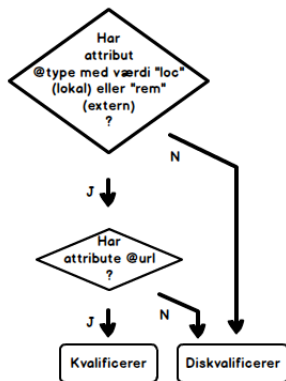
Beslutning: //oa_link element kvalificerer?

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

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

At checke kvalificeringen gøres derfor med følgende lille workflow:

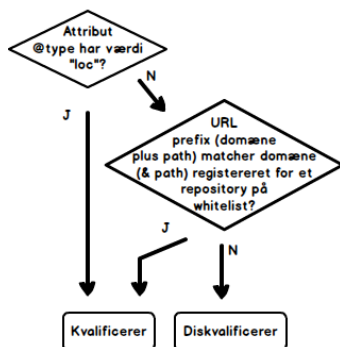
//oa_link element kvalificerer?



Beslutning: 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 prefix (domænenavn – muligvis sammensat med path) registreret for et repositorie på autoritetslisten for accepterede eksterne repositorer ("Whitelist"). At checke kvalificeringen gøres derfor med følgende lille workflow:

URL kvalificerer?



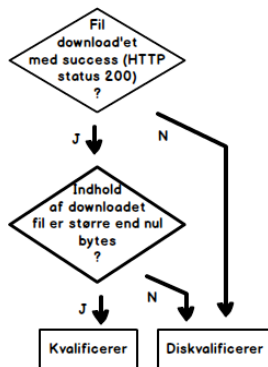
Beslutning: Fil kvalificerer?

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?



4.1.3 Check for Uudnyttet og Uklart Potentiale

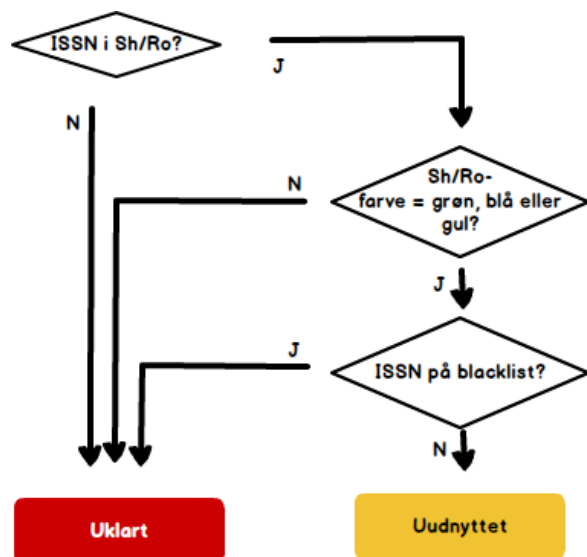
Hvis metadatatypen ikke har et realiseret Open Access Potentiale, undersøges, om potentialet er Uudnyttet. Er det ikke uudnyttet, defineres potentialet som Uklart.

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>).

Anvendt regel:

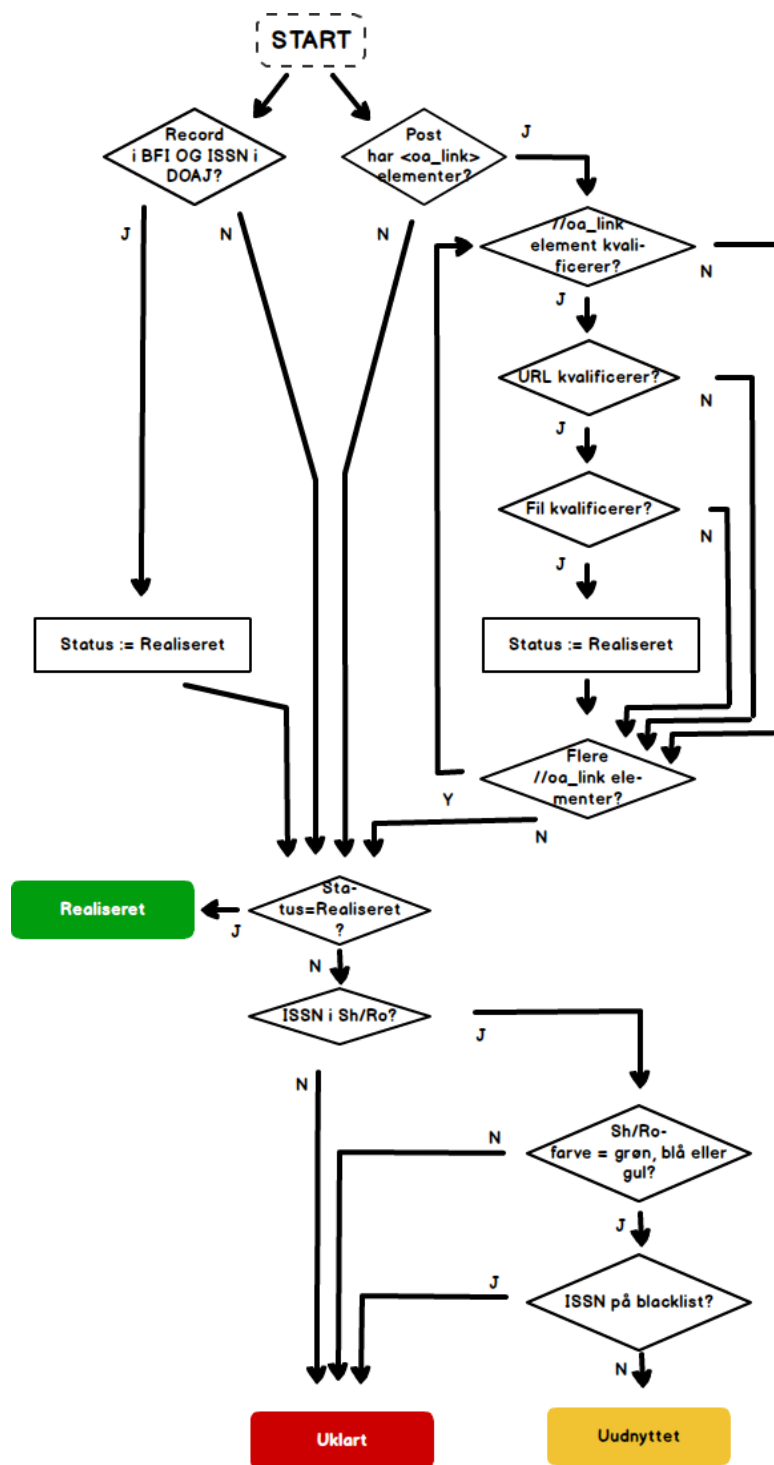
- Er tidsskriftets ISSN nummer noteret i Sherpa/Romeo med farvekoderne grøn, blå eller gul, har tidsskriftet OA potentiale, og publikationen anses for at have **Uudnyttet** Open Access potentiale.
 - En undtagelse til denne regel er, hvis tidsskriftets ISSN nummer er noteret for et tidsskrift på Autoritetslisten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

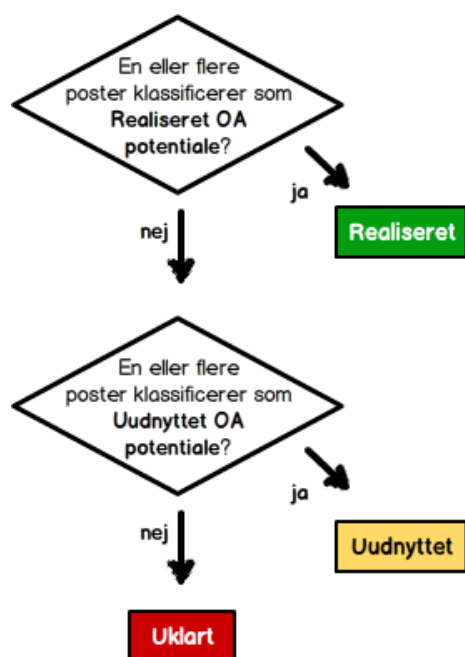


4.2 Open Access klassifikation – på nationalt/hovedforskningsområde niveau

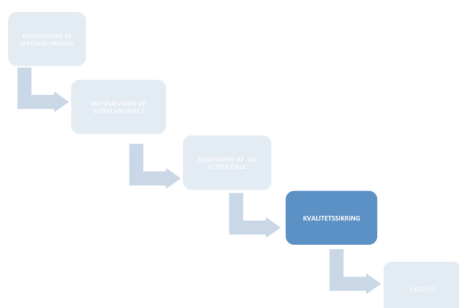
Publikationsposter i genstandsfeltet uden dubletter svarer til klynger hver bestående af een eller flere poster fra genstandsfeltet med dubletter.

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:



5 Proces 4: Kvalitetssikring



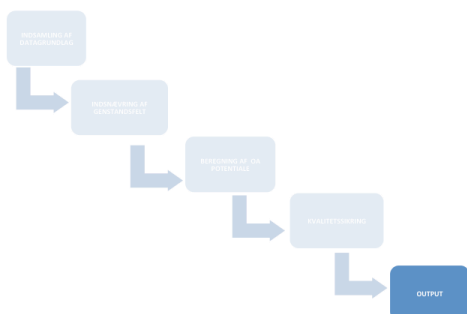
Open Access Indikatorens resultater har været udsat for nedenstående kvalitetssikringsprocedurer:

- **Datagrundlaget.** De indsamlede data og de registrerede links til fulltekster og tilstedeværelsen i universiteternes forskningsdatabaser af de filer, som links peger på, er blevet testet. Testen har været stikprøve baseret og har været repræsentativ på tværs af de indgående universiteter.
- **Downloadede fuldtekstfiler.** De indsamlede data og de registrerede links samt tilstedeværelsen af de filer der linkes til i universiteternes forskningsdatabaser, er blevet testet. Et udvalg af de downloadede fuldtekstfiler er blevet manuelt undersøgt for at sikre, at de reelt repræsenterer den videnskabelige artikel, som publikationsposten beskriver – på en komplet og læsbar måde. Testen har fokuseret

på downloadede filer, der ved simple computerbaserede analyser synes at afvige fra de registrerede metadata. Dette har f.eks. drejet sig om afvigelser i sidetal, meget små filstørrelser mv.

- **Links til eksterne repositorier.** Alle filer der er realiseret via links til anerkendte eksterne OA repositorier er gennemgået med henblik på at sikre, at de registrerede links leder til fuldtekstfiler som repræsenterer den videnskabelige artikel.
- **Stikprøve.** En tilfældig udvalgt stikprøve på 5%, af det samlede datasæt på realiseret Open Access potentiale fra hvert universitet, er gennemgået med henblik på et overordnet kvalitetstjek af datakvaliteten.

6 Proces 5: Output



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

Den Danske Forskningsdatabase (<http://forskningsdatabasen.dk/>) anvendes som formidlingsplatform for visualiseringer såvel som for rapporterne.

6.1 Datarapporter til download

Der produceres fem datarapporter:

- 1) Summeringer: Genstandsfelterne optalt samlet og fordelt på **Realiseret**, **Uudnyttet** og **Uklart** 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 på Den Danske Forskningsdatabases hjemmeside

Open Access Indikatorens summeringer visualiseres på

http://forskningsdatabasen.dk/da/open_access/overview, hvor også datarapporter kan downloades.

7 Appendix A: Fulltext Download delprocessen

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

Denne delproces er implementeret på følgende måde:

- Fulltexter downloades en ad gangen (serielt; ikke i parallelt)
- Fulltexter downloades fra universiteterne på skift:
 - en fulltext fra universitet 1
 - en fulltext fra universitet 2
 - en fulltext fra universitet 3
 - ...
 - en fulltext fra universitet N
 - en fulltext fra universitet 1
 - en fulltext fra universitet 2
 - ...
 - en fulltext fra universitet N
 - ...
 - ...

Alle downloads foretages automatiseret af OA Indikatorens download robot.

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

- IP Adresse: 192.38.67.38
-