

KPI 4: Open Data

Open Science

Uitwerking van de indicator in FRIS

In deze fiche wordt beschreven hoe de KPI4 "Open Data", zoals gedefinieerd in de KPI-nota van de FOSB van 4/12/2020 (zie Referenties) concreet wordt gemeten op basis van de gegevens in FRIS. Volgens de KPI-nota zou de eerste nulmeting plaatsvinden in 2021 voor elk van de instellingen die vallen onder het Open Science initiatief van de Vlaamse overheid (de universiteiten, de hogescholen, de instellingen voor postinitieel onderwijs, de Vlaamse wetenschappelijke instellingen, de strategische onderzoekscentra, en overige centra). De uitvoering van die nulmeting werd later met een jaar opgeschoven om de instellingen de tijd te geven zich te organiseren om de data te verzamelen. Vanaf 2022 zal de meting wel maximaal gebeuren via FRIS. Onderstaande fiche beschrijft hoe die meting concreet zal gebeuren.

Voor de instellingen die (nog) niet aanleveren aan FRIS, wordt de KPI door hen zelf gemeten en dit maximaal conform de binnen Open Science afgesproken meetmethode.

Uit de KPI-nota van de FOSB (zie Referenties):

Wat:

a. de digitale onderzoeksdata voortkomend uit onderzoek (geheel of gedeeltelijk) gefinancierd met Vlaamse publieke middelen en die aan de basis liggen van peer-reviewed tijdschriftartikels. De hier voorgestelde KPI heeft geen betrekking op 'levende' onderzoeksdata, enkel op onderzoeksdata die gearchiveerd zijn en dus niet verder bewerkt worden.

b. 'Open Data' zijn digitale onderzoeksdata, volgens bovenstaande definitie, die voor iedereen toegankelijk zijn en die door iedereen gebruikt, gewijzigd en/of gedeeld kunnen worden en dit voor elk doeleinde. Een volledige definitie impliceert het voorkomen van expliciet gemelde open licenties en het gebruik van open data-formaten. Gezien dit voor de meeste disciplines nog nieuwe materie is en voor bepaalde types onderzoeksdata geen open data-formaten voorhanden zijn, stelt de FOSB WG RDM & OS voor om in twee stappen te werken:

a. de data zijn voor iedereen toegankelijk (eenvoudige definitie, te gebruiken tot 2022)

b. de data zijn voor iedereen toegankelijk en gaan gepaard met een (open) licentie (uitgebreide definitie, te gebruiken vanaf 2023)

Einddoel:

De KPI stelt dat voor 100% van de peer-reviewed wetenschappelijke artikels met onderliggende digitale onderzoeksdata, voortkomend uit onderzoek (geheel of gedeeltelijk) gefinancierd met Vlaamse publieke middelen, zijn deze onderliggende digitale onderzoeksdata beschikbaar als Open Data, al dan niet na afloop van een embargo, voor zover geen legitieme opt-out voor Open Data geldt¹. *Dit doel dient, afhankelijk van de nulmeting, bereikt te worden in 2024, 2025 of 2026.*

Definitie KPI:

i. populatie: in FRIS geregistreerde metadatabeschrijvingen van datasets voortkomend uit onderzoek (geheel of gedeeltelijk) gefinancierd met Vlaamse publieke middelen en die aan de basis liggen van peer-reviewed wetenschappelijke artikels;

¹ Deze definitie is analoog aan deze recent geformuleerd in <https://www.cesaer.org/content/5-operations/2020/20200610-white-next-generation-metrics.pdf>.

ii. meting:

i. noemer: alle peer-reviewed tijdschriftartikelen met onderliggende dataset(s) voortkomend uit onderzoek (geheel of gedeeltelijk) gefinancierd met Vlaamse publieke middelen met uitzondering van artikelen waarvoor legitieme opt-outs de publicatie van Open Data verhinderen;

ii. teller eenvoudige definitie: aandeel in bovenstaande tijdschriftartikelen van tijdschriftartikelen met minstens één dataset die open beschikbaar is volgens de eenvoudige definitie van Open Data;

iii. teller uitgebreide definitie: aandeel in bovenstaande tijdschriftartikelen met minstens één dataset die open beschikbaar is volgens de uitgebreide definitie van Open Data, d.i. eveneens voorzien van een open licentie;

iv. Deze KPI veronderstelt een formele koppeling (in FRIS) tussen de dataset en de publicatie daar waar de publicatie het resultaat is van een Vlaamse financiering én een gedifferentieerd Open Data label in FRIS, dat bestaat uit twee componenten: enerzijds een aanduiding van het toegangsniveau van de dataset, waarbij minstens onderscheid maakt tussen 'open', 'onder embargo' (met vermelding van expiry date) en 'om legitieme redenen beperkt/restricted'² of 'om legitieme redenen gesloten', anderzijds een indicatie van het type licentie of ruimer de eventuele intellectuele eigendomsrechten verbonden aan de dataset. De eerste component volstaat voor de meting van de KPI volgens de eenvoudige definitie; de combinatie van beide componenten laat toe te meten volgens de uitgebreide definitie.

v. De kennisinstellingen zijn verantwoordelijk voor de correcte invulling en aanlevering van beide componenten van dit label (inclusief wijziging van status 'onder embargo' naar 'open' na aflopen van embargoperiode indien die niet automatisch gebeurt);

iii. Bij ontbreken van (automatische) aanlevering aan FRIS, wordt dit gemonitord via het jaarlijks verslag van alle kennisinstellingen.

Verdere uitwerking in FOSB werkgroep Metadata en Standaardisatie en implementatie in FRIS:**Uitwerking metadatamodel en definities:**

In de werkgroep Metadata en Standaardisatie werd het metadatamodel voor datasets uitgewerkt (zie Referenties). Dit datamodel bevat tevens de nodige attributen om de Open Data KPI te kunnen meten. Het document bevat daarnaast ook definities van de gebruikte begrippen, waaronder die van Datasets:

"Dataset: Data and objects generated or collected by researchers affiliated by a Flemish research institution in the course of their investigations, regardless of their form or method, that form the object on which researchers test a hypothesis. This includes the full range of data: raw, unprocessed datasets, proprietary generated and processed data and secondary data obtained from third parties."

Implementatie in FRIS:

Het metadatamodel werd vervolgens ontwikkeld in het FRIS-platform.

Meer informatie over het volledige FRIS-datamodel, de gebruikte attributen en de afgesproken businessrules is te vinden in het FRIS Vademecum (zie Referenties).

² Volgens de definities gehanteerd door SURF, cf. <https://wiki.surfnet.nl/display/standards/info-eurepo#infoeurepo-AccessRights>.

Meting van de KPI in FRIS:	
Populatie:	
Voorwaarden volgens KPI-nota:	Uitwerking in FRIS:
in FRIS geregistreeerde metadatabeschrijvingen van datasets	Publicaties met een link naar een dataset in FRIS
die aan de basis liggen van peer-reviewed wetenschappelijke artikels	<p>Publicaties van</p> <ul style="list-style-type: none"> - het Research Output Type = "Journal Contribution" en = één van de volgende subtypes: "Journal Article" (of "e-pub"), "Journal Article Conference Contribution", "Letter", "Research Note" of "Review Article" - én met status "published" - én met een Tijdschrift ISSN dat voorkomt op één of meerdere van de volgende lijsten (zie Referenties): <ul style="list-style-type: none"> • de lijst van de Web of Science Journals zoals gebruikt door ECOOM • de cumulatieve lijst van peer-reviewed Journals van VABB (Vlaams Academisch Bibliografisch Bestand voor de Sociale en Humane Wetenschappen) • de Master Journal List van Clarivate (Web of Science) • de Journals van DOAJ
voortkomend uit onderzoek (geheel of gedeeltelijk) gefinancierd met Vlaamse publieke middelen (Dit veronderstelt een formele koppeling (in FRIS) tussen de dataset en de publicatie daar waar de publicatie het resultaat is van een Vlaamse financiering)	Publicaties gelinkt aan één of meerdere projecten met een fincode die verwijst naar een Vlaamse financieringsbron . De lijst met fincodes van Vlaamse financieringsbronnen is deze zoals ter beschikking gesteld door ECOOM Hasselt (FRISfincodes-Vlaamse financiering en DMP-compliance, zie Referenties).
Publicaties die aan bovenvermelde voorwaarden voldoen, worden meegenomen in de populatie.	

Meting:	
Voorwaarden volgens KPI-nota:	Uitwerking in FRIS:
<p>noemer: alle peer-reviewed tijdschriftartikelen met onderliggende dataset(s) voortkomend uit onderzoek (geheel of gedeeltelijk) gefinancierd met Vlaamse publieke middelen met uitzondering van artikelen waarvoor legitieme opt-outs de publicatie van Open Data verhinderen;</p>	<p>De publicaties uit de populatie van hierboven gekoppeld aan een dataset voor zover die geen legitieme opt-out heeft. Dit wil zeggen dat datasets waarvoor een opt-out is aangeleverd, worden uitgesloten van de meting.</p>
<p>teller eenvoudige definitie (te gebruiken tot 2023³): aandeel in bovenstaande tijdschriftartikelen van tijdschriftartikelen met minstens één dataset die open beschikbaar is volgens de eenvoudige definitie van Open Data;</p>	<p>Het aandeel van de publicaties uit de populatie van hierboven gekoppeld aan een dataset waarvoor het veld 'Toegangsrechten' de volgende waarden heeft: 'open' of 'onder embargo'. De deelpercentages voor Open en voor onder embargo zullen ook ter informatie worden meegegeven.</p>
<p>teller uitgebreide definitie (te gebruiken vanaf 2024⁴): aandeel in bovenstaande tijdschriftartikelen met minstens één dataset die open beschikbaar is volgens de uitgebreide definitie van Open Data, d.i. eveneens voorzien van een open licentie;</p>	<p>De teller van de eenvoudige definitie met bijkomende voorwaarde dat er aan de publicatie een dataset is gekoppeld met een licentie die open is. Het open karakter van de licentie die hoort bij de dataset, wordt op het moment van de KPI-meting afgeleid uit de informatie zoals beschikbaar bij opendefinition.org en de SPDX-identificer van de licentie bij de dataset zoals aangeleverd door de instelling (zie Referenties).</p>

Tijdsdimensie:
<p>Jaar:</p> <p>Het jaar is het jaar waarop de meting betrekking heeft. Voor het jaar wordt gebruik gemaakt van het <u>publicatiejaar van de publicatie</u> waar de dataset mee gelinkt is en die een link heeft naar een project met Vlaamse financiering. Voor de meting worden noch het publicatiejaar van de dataset zelf, noch de tijdsdimensie van het gekoppelde project met Vlaamse financiering in rekening gebracht.</p> <p><i>Bv. het cijfer over het jaar 2021 wordt afgeleid uit de publicaties met publicatiejaar 2021 die een gekoppelde dataset hebben en die een link hebben naar een project met Vlaamse financiering.</i></p>
<p>Nulmeting en groeitrajecten voor de KPI:</p> <p>De nulmeting heeft betrekking op het jaar 2021. In de KPI-nota (zie Referenties) was sprake van 2020, maar er werd voor deze KPI een jaar uitstel gevraagd en gekregen.</p> <p>Afhankelijk van de waarde van de nulmeting is een verschillend groeitraject met de bereiken doelpercentages voor de KPI van toepassing. Er zijn 3 verschillende groeitrajecten mogelijk (cfr. KPI-nota, zie Referenties).</p>

³ In de KPI nota stond tot 2022 voor de eenvoudige definitie, maar omdat de nulmeting een jaar is uitgesteld, is ook deze datum een jaar opgeschoven.

⁴ In de KPI nota stond vanaf 2023 voor de uitgebreide definitie, maar omdat de nulmeting een jaar is uitgesteld, is ook deze datum een jaar opgeschoven.

Meetjaar en timing rapportering:

Het meetjaar is het jaar waarin de meting gebeurt. De eerste meting gebeurt eind april van het meetjaar.

De meetgegevens worden aan de CPOC van de instelling bezorgd samen met de achterliggende details. De instelling neemt dit cijfer op in haar Open Science jaarrapport dat tegen 15 juni moet worden ingediend. De instelling kan het meetcijfer duiden in het jaarverslag, bv. wanneer het om bepaalde redenen afwijkt van de realiteit. Als er gegevens ontbreken in de meting, kunnen die enkel worden aangevuld via FRIS zelf.

Indien gewenst kan de instelling nog tot eind mei om een update van de meting in FRIS vragen via fris@vlaanderen.be.

De meting gebeurt jaarlijks over het voorafgaande jaar. Tevens worden er updates gegeven van de metingen van de voorafgaande jaren teruggaand tot en met de nulmeting.

Bv. in 2024 wordt gemeten over jaar 2023 en wordt een update gegeven van de cijfers over de jaren 2022 en 2021 (nulmeting).

Attributen in FRIS, gebruikt voor de KPI-meting:

Attribuut	Object of classificatie	Commentaar
publicatie-ID	RESEARCH OUTPUT	wordt gebruikt om aantallen te tellen
Publicatietype	Classificatie op object RESEARCH OUTPUT	om te filteren op relevante Journal contribution types
Journal ISSN	RESEARCH OUTPUT	om te filteren op de lijsten van Journals die peer reviewed zijn
Publicatiejaar	RESEARCH OUTPUT	voor tijdsaspect van de meting
link naar project	link PROJ-RESEARCH OUTPUT	om te filteren op projecten gefinancierd met Vlaamse publieke middelen
Fancode	PROJ FUNDING	
link naar dataset	link RESEARCH OUTPUT- DATASET	om die publicaties te filteren waar datasets aan gelinkt zijn
Toegangsrechten	DATASET	om te filteren op "Open" of "embargoed"
Opt-out	DATASET	Om de datasets met legitieme opt-out uit te filteren
Licentie (SPDX short identifier)	DATASET	om vanaf 2023 te kunnen filteren op licenties die open zijn. Het kenmerk "open" zal op het moment van de KPI-meting worden opgehaald bij Opendefinition.org o.b.v. de SPDX-identifier van de licentie (aangeleverd door de instelling).

Meer informatie over het volledige FRIS-datamodel, de gebruikte attributen en de afgesproken businessrules is te vinden in het FRIS Vademecum (zie Referenties).

Referenties		
Lijst of document	Omschrijving	Link
KPI-nota	Mededeling aan de Vlaamse Regering van 4/12/2020 waarin de KPI's voor Open Science zijn gedefinieerd.	https://www.ewi-vlaanderen.be/sites/default/files/bestanden/5fc5f512b328e9000c0007f3.pdf
Metadatamodel voor datasets	Het metadatamodel voor datasets zoals uitgewerkt door de FOSB werkgroep Metadata en Standaardisatie (2020-11-20_FOSB_metadatamodel_V1.7ter_FRIS implementation guidelines v2)	https://researchportal.be/nl/open-science
FRIS Vademecum	Naslagwerk waar informatie over het FRIS-datamodel, de gebruikte attributen en de afgesproken businessrules op een overzichtelijke manier zijn samengebracht (bedoeld voor businessmensen).	https://researchportal.be/nl/open-science
FRISfincodes-Vlaamse financiering en DMP-compliance	Dit is de lijst met fincodes van Vlaamse financieringsbronnen en wordt jaarlijks geupdated	https://researchportal.be/nl/open-science
de lijst van de Web of Science Journals zoals gebruikt door ECOOM	Deze lijst bevat tijdschriften uit volgende collecties uitgegeven door Clarivate Analytics: Science Citation Index (SCIE) of Social Science Citation Index (SSCI) met en zonder een impactfactor, Arts and Humanities Citation Index (AHCI), Science & Technology Database- en Social Sciences & Humanities Database-proceedings	https://researchportal.be/nl/open-science
Master Journal List van Clarivate (Web of Science)	Dit is de lijst van tijdschriften, gebruikt in de Web of Science Core Collection™: the Science Citation Index Expanded™ (SCIE), Social Sciences Citation Index™ (SSCI), Arts & Humanities Citation Index™ (AHCI), and Emerging Sources Citation Index™ (ESCI)	https://mjl.clarivate.com/collection-list-downloads
Peer-reviewed Journals van VABB	Dit is de cumulatieve lijst van peer-reviewed Journals van het Vlaams Academisch Bibliografisch Bestand voor de Sociale en Humane Wetenschappen (VABB)	https://researchportal.be/nl/open-science
DOAJ Journal list	Dit is de lijst van Open Access tijdschriften zoals opgelijst in de Directory of Open Access Journals	https://doaj.org/csv
Licenties van SPDX.org	Lijst met licenties gebruikt in software, data, hardware en documentatie	https://SPDX.org/licenses/
Lijst van Opendefinition.org	Lijst met informatie over het open karakter van een licentie	https://opendefinition.org/licenses/