



Adviesrapport

Metadata standaarden en afspraken voor praktijkgericht onderzoek

**Auteurs**Maarten Hoogerwerf Marta Kargól Tineke van der Meer Renate Mattisziik Roel Oomens Eileen Waegemaekers **Datum**

9 juli 2024

Licentie

CC-BY 4.0

Versie

1.0

DOIdoi.org/10.5281/zenodo.12701231

Inhoudsopgave

1. Inleiding en doel van dit rapport	3
Aanleiding	3
Aanpak werkgroep	5
2. Algemene verkenning	6
Nationale ontwikkelingen	6
Internationale ontwikkelingen	7
Informatiebehoefte en waardeketen	7
Informatiemodellen en afsprakenstelsels	9
Interoperabiliteit	10
Commons, of gemeenschapsgoederen	10
3. Use case: Publinova	12
Wat is Publinova?	12
Informatiebehoefte en waardeketen	12
Informatiemodel en afsprakenstelsel	13
Persistent identifiers	14
Commons	15
4. Bevindingen en aanbevelingen	17
Bijlage 1 Tabel Entiteit Persoon	20
Bijlage 2 Tabel Entiteit Project	21
Bijlage 3 Tabel Entiteit Partij	22
Bijlage 4 Tabel Entiteit Product	23

1. Inleiding en doel van dit rapport

Praktijkgericht onderzoek heeft zich de afgelopen jaren sterk ontwikkeld en speelt een bijzondere rol in het Nederlandse kennissysteem¹. Na de sterke groei van het aantal lectoren ligt de focus nu op het bouwen van het fundament dat praktijkgericht onderzoek ondersteunt en verder professionaliseert. Om dit fundament te bouwen is onder andere inzicht en overzicht nodig van het praktijkgericht onderzoek.

Om praktijkgericht onderzoek te kunnen profileren en hergebruiken, maar ook om te kunnen sturen en verantwoorden, hebben we onderzoeksinformatie nodig. Onderzoeksinformatie betreft informatie **over** onderzoek, zoals de onderzoekers werkzaam aan de hogeschool, onderzoeksthema's, projectinformatie, samenwerkingspartners, kennisproducten etc. Als we onderzoeksinformatie willen uitwisselen en hergebruiken, dan zijn daar eenduidige afspraken en standaarden voor nodig. Zonder afspraken over de standaarden die we gebruiken ontstaan er interpretatieverschillen en kan de informatieketen niet goed functioneren. Het ontwikkelen, beheren en implementeren van zulke afspraken en standaarden vergt actieve participatie van de community, commitment van organisaties op het volgen van deze afspraken, en expertise van kennisinstellingen om de benodigde ondersteuning te bieden.

Dit adviesrapport geeft een overzicht van de informatiebehoeften en de bijbehorende standaarden. Op basis van de geleerde lessen geven we aanbevelingen hoe we als sector toe kunnen werken naar een standaard voor uitwisseling van onderzoeksinformatie.

Aanleiding

In juni 2023 is vanuit het project 'Onderzoeksinformatie in het hbo' bij SURF de werkgroep metadata gestart met als doel om afspraken te maken over de standaarden die we gebruiken om onderzoeksinformatie uit te wisselen. Er ligt een duidelijke behoefte om informatie op een gestandaardiseerde manier te registreren en uit te wisselen. Deze behoefte kwam via meerdere wegen bij SURF terecht.

In 2020 voerde SURF een verkenning uit naar hun behoeftes op het vlak van onderzoeksinformatie bij de hogescholen. Hierbij werd rekening gehouden met de diversiteit die het PO rijk is. In 20 interviews bij 8 hogescholen is er gesproken met informatiespecialisten, onderzoekers, subsidieadviseurs en beleidsadviseurs. De resultaten van de interviews zijn verwerkt tot een [adviesrapport](#) (juli 2020)² met daarin een aantal aanbevelingen.

¹ Rathenau Instituut (2024). *Monitor praktijkgericht onderzoek 2022*. <https://www.rathenau.nl/nl/werking-van-het-wetenschapssysteem/monitor-praktijkgericht-onderzoek-2022>

² Pijnacker, J., Teirlinck, M. & Mahler, R. (2020). *Adviesrapport Verkenning Onderzoeksinformatie in het hbo*. SURF. <https://doi.org/10.48544/b5468f50-315d-4646-845a-d73e614f23e2>

Aanbevelingen

- Hogescholen bevinden zich op het gebied van onderzoeksinformatie in verschillende fases. Gebruik daarom de scenario's als een **groeimodel** met verschillende instapniveaus voor hogescholen.
- Zoek naar een gezamenlijke basis voor de registratie van een **minimale set** van onderzoeksinformatie
- Focus op de **uitwisseling** van informatie tussen hogescholen, niet op een systeem. Hierbij gaat het om afspraken over welke informatie men uitwisselt en de (inter)nationale standaarden waarmee dat gebeurt.

Alhoewel in dit rapport de nadruk al werd gelegd op uitwisselingsstandaarden, is aan dit advies geen gevolg gegeven. Er wordt veel kennis gedeeld maar op bovenstaande aanbevelingen is nog weinig gezamenlijke actie gekomen.

De komst van **Publinova** zet onderzoeksinformatie nog meer in de schijnwerpers en maakt de hoge urgentie bij de instellingen om informatie te registreren en uit te wisselen zichtbaar. Er is bij de instellingen veel beweging om oplossingen te vinden of te ontwikkelen om onderzoeksinformatie te registreren. Ook vanuit het DCC-PO³ is de vraag gekomen om de randvoorwaarden te realiseren voor het ontwikkelen van een modulair Research Information System (RIS), waarmee bedoeld wordt op de afspraken die nodig zijn om de onderzoeksinformatie op een gestandaardiseerde manier te registreren en uit te wisselen.

Parallel is bij SURF het programmaplan Open Research Information opgesteld met de volgende doelstelling:

“Alle informatie over Nederlands publiek gefinancierd onderzoek en de resultaten daarvan is open beschikbaar en herbruikbaar.”

– SURF ledenraad, 13 dec 2023

In dit programmaplan wordt ook digitale soevereiniteit genoemd als kernwaarde voor open onderzoeksinformatie. Onderzoeksinformatie is een gemeenschapsgoed dat we gezamenlijk moeten onderhouden op basis van open standaarden en publieke

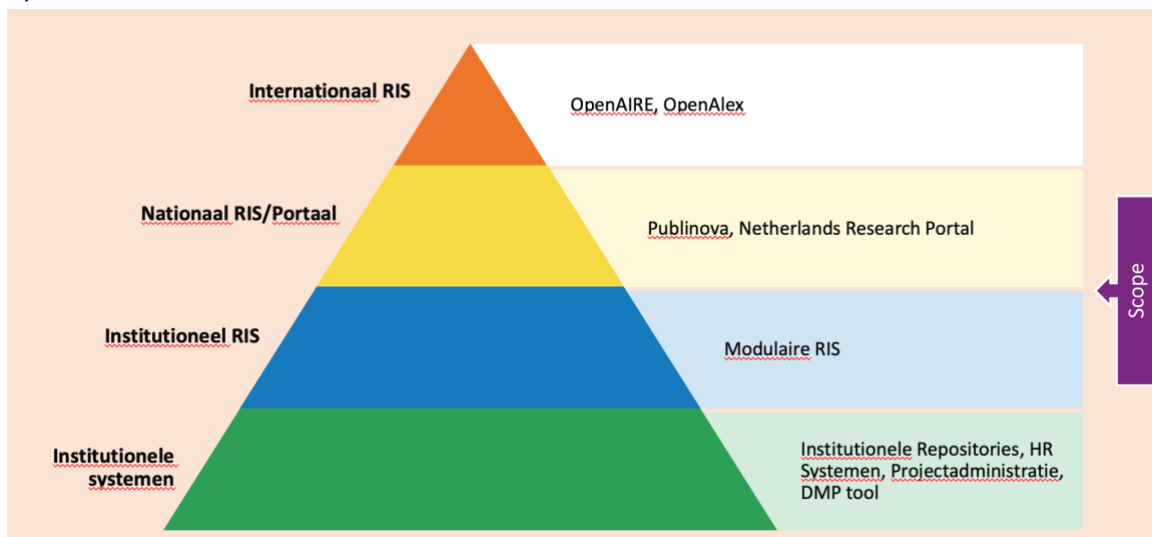
³ Digital Competence Center voor Praktijkgericht Onderzoek, <https://dcc-po.nl/>

governance. Het bereiken van dit doel vergt een complexe transformatie van het huidige systeem en vraagt om “collective action”.

Aanpak werkgroep

Zoals in de vorige paragraaf is beschreven, zijn er meerdere aanleidingen geweest om in actie te komen. Bij aanvang van de werkgroep is onderkend dat de informatiebehoefte groot en divers is en de werkgroep is dan ook gestart om de omvang in te kunnen schatten en de opdracht aan te scherpen.

Hoewel de standaarden en afspraken die nodig zijn voor de informatieketen betrekking hebben op instellings-, nationaal, en internationaal niveau oriënteren we ons als werkgroep op instellingsoverstijgende uitwisseling van onderzoeksinformatie (zie Figuur 1).



Figuur 1 Overzicht van de verschillende lagen in de onderzoeksinformatieketen

We kijken daarbij in de volle breedte naar de informatiebehoefte, waardeketen, en huidige standaarden. Om focus en waarde te geven aan de oriëntatie gebruiken we Publinova (het nationaal platform voor het vindbaar en zichtbaar maken van praktijkgericht onderzoek) als use case. Met de selecte set aan onderzoeksinformatie die Publinova vraagt kunnen we de resultaten tastbaar maken en de inzichten toepassen. Als aanpak is gekozen om agile en waardegericht te werken, waarbij we continu bouwen aan kennis en communityvorming, met een focus op uitwisseling van onderzoeksinformatie van de instellingen naar Publinova.

Naast het toewerken naar een standaard heeft de werkgroep ook de taak om bewustzijn te creëren voor het werken met een standaard en het draagvlak voor te bereiden. Dit doen we door diverse stakeholders te betrekken (van onderzoekers tot bestuurders) en onze opgedane kennis te delen op community events.

2. Algemene verkenning

In deze sectie verkennen we het domein van onderzoeksinformatie in praktijkgericht onderzoek. Deze verkenning richt zich op de volle breedte, maar we laten de verkenning wel leiden door vraagstukken vanuit onze use case. We doen dit aan de hand van de volgende vragen:

- Wat zijn waardeketens rondom de onderzoeksinformatie in het hbo?
- Wat zijn de standaarden en afsprakenstelsels rondom het beheren en uitwisselen van onderzoeksinformatie?

We onderzoeken dit vanuit de geïdentificeerde ontwikkelingen binnen en buiten Nederland en nemen de vragen vanuit de use case hierin mee. Het doel is in deze fase niet om tot een uitvoerig of compleet overzicht te komen, maar om de inzichten te borgen die we bij de verschillende activiteiten hebben opgedaan.

Nationale ontwikkelingen

Op nationaal niveau is SURF betrokken bij de volgende ontwikkelingen:

- **Nationale Repository en (C)RIS infrastructuur.** In het praktijkgericht onderzoek zijn de belangrijkste diensten SURFsharekit als repository, HBO Kennisbank als gefedereerd portaal over repositories, en Publinova voor profilering. In het wo wordt veel gebruik gemaakt van Pure maar worden ook andere systemen gebruikt. Nationaal was er tot voor kort NARCIS, en is nu het Netherlands Research Portal opgezet in samenwerking met OpenAIRE⁴.
- **PID Infrastructuur.** SURF faciliteert het Nederlandse ORCID Consortium⁵ met op dit moment 18 leden. Daarnaast zijn er Persistent Identifier (PID) diensten zoals Handle, RAiD (in ontwikkeling) en de URN:NBN Resolver.
- **PID Roadmap.** Een initiatief van NWO, DANS-KNAW, UKB, SURF en CWTS om de initiatieven op het gebied van toekennen en gebruik van persistent identifiers te coördineren⁶.
- **PID graph.** Een pilot onder de Open Research Information Agenda (ORIA) om inzicht te krijgen in de waardepropositie en technische requirements voor een PID Graph, inzicht te krijgen in de datakwaliteit van onderliggende bronnen, en bij te dragen aan de communityvorming rondom open onderzoeksinformatie.⁷

⁴ <https://netherlands.openaire.eu/>

⁵ <https://www.surf.nl/orcid>

⁶ Ackaova, G. et al (2022). *Towards a national PID roadmap*. <https://doi.org/10.5281/zenodo.7660190>

⁷ Sweeney, J., Waegemaekers, E., & Hoogerwerf, M. (2024). National PID graph Pilot Report (ORIA). Zenodo. <https://doi.org/10.5281/zenodo.10610929>

- **WISH werkgroep.** Deze Edustandaard werkgroep richt zich op onderzoeksmetadata van en voor het hoger onderwijs en onderhoudt de volgende afspraken: Semantics, DIDL, MODS, OAI-PMH, URN.⁸
- **Overig.** Er loopt een kleinschalig onderzoek bij een aantal hogescholen naar de relatie tussen een uitwisselingsstandaard en de interne processen voor het verzamelen en delen van onderzoeksinformatie.

Internationale ontwikkelingen

Op internationaal niveau zien we de volgende ontwikkelingen:

- **EuroCRIS.** Via SURF is Nederland lid van [euroCRIS](#), de organisatie die [CERIF](#) beheert en doorontwikkelt. Via euroCRIS zijn we in contact met het Finse [research.fi](#)⁹, en de bedenkers van het Belgische DOSP, een projectadministratiesysteem voor de hogescholen dat gegevens uitwisselt naar het Vlaamse portaal [FRIS](#)¹⁰.
- **OpenAIRE.** Nederlandse instellingen zijn altijd voorloper geweest in de ontwikkeling en het leveren van informatie aan [OpenAIRE](#), een organisatie die diensten en infrastructuur levert om Open Science, Open Data en Open Access te bevorderen. Met het wegvallen van NARCIS hebben we via OpenAIRE snel het Netherlands Research Portal kunnen realiseren.
- **Research Data Alliance (RDA).** In [RDA](#) wordt gewerkt aan o.a. [SKG-IF](#)¹¹, een interoperability framework voor Science Knowledge Graphs (zoals de OpenAIRE Graph)
- **EOSC.** SURF neemt deel aan [FAIRCORE4EOSC](#). Hierin voeren we een case study uit voor het verbinden van nationale PID Graphs. Onderdeel is om samen met euroCRIS een nieuwe versie van CERIF te ontwikkelen. Daarnaast zet SURF i.s.m. EOSC een Europees Registration Agency op voor RAiD, een PID service voor onderzoeksprojecten
- **Knowledge Exchange.** SURF is één van de 6 Europese partners in deze [expert groep](#) waarin we samenwerken om de ontwikkeling van digitale infrastructuren die Open Science mogelijk maken te ondersteunen.

Informatiebehoefte en waardeketen

Binnen het DCC-PO, maar ook bij het ontwikkelen van de nationale PID Roadmap wordt sterk gehecht aan een waardeketen. Om die reden is tijdens het project PID graph, en ook binnen dit project onderzoeksinformatie in het hbo gewerkt met een waardeketen. Op deze manier krijgen we grip op de verschillende rollen die onderzoeksinformatie

⁸ https://www.edustandaard.nl/standaard_werkgroepen/werkgroep-metadata-subwerkgroep-wish/

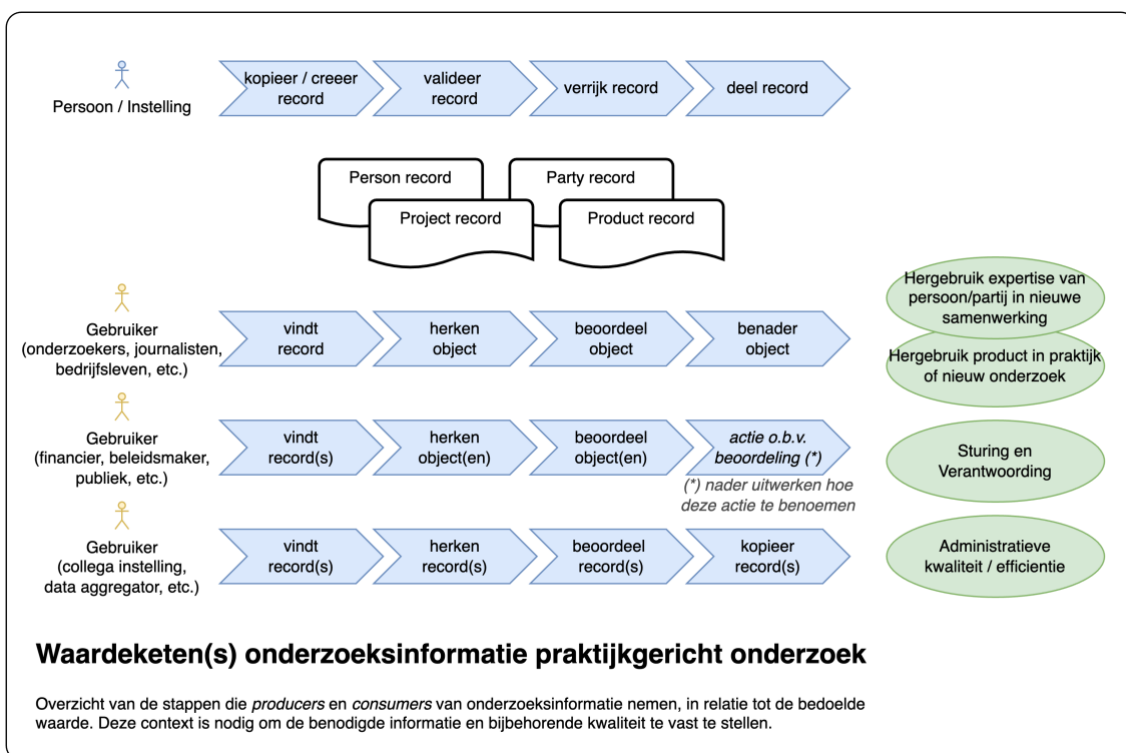
⁹ <https://research.fi/en/>

¹⁰ <https://researchportal.be/nl>

¹¹ Scientific Knowledge Graphs - Interoperability Framework, <https://skg-if.readthedocs.io/en/latest/>

speelt in de informatiebehoefte van de sector. In onderstaande diagram (Figuur 2) zien we globaal drie toepassingen voor het registreren en beheren van onderzoeksinformatie:

1. **Profileren en hergebruik.** Onderzoekers en instellingen profileren zichzelf en hun werk, zodat hun kennis en producten gevonden, gewaardeerd, hergebruikt en/of doorontwikkeld worden in nieuwe samenwerkingen of projecten.
2. **Sturing en verantwoording.** Onderzoekers en instellingen geven inzicht in hun onderzoek en resultaten, via rapportages aan managers en andere stakeholders of via een portaal aan externe partijen. Hiermee wordt weer richting gegeven aan nieuw uit te voeren onderzoek. Een veelgehoorde wens is ook inzicht in lopend onderzoek, zodat kan worden gestuurd op nieuw uit te voeren onderzoek.
3. **Administratieve efficiëntie en kwaliteit.** Hoewel dit eigenlijk onderdeel is van de bovengenoemde waardeketens, is het een belangrijke reden voor afspraken voor het registreren en ontsluiten van onderzoeksinformatie.



Figuur 2: Schets van de globale waardeketens voor onderzoeksinformatie

We zien dat de informatiebehoefte in praktijkgericht onderzoek verbreedt van onderzoeksoutput (zoals ontsloten via SURFsharekit en HBO Kennisbank) naar personen, projecten en partijen (zoals geprofileerd in Publinova). Wanneer de informatie wordt gebruikt voor sturing en verantwoording, zullen ook aanvullende afspraken nodig zijn om de kwaliteit van de onderzoeksinformatie te borgen.

Informatiemodellen en afsprakenstelsels

In Tabel 1 hieronder worden de verschillende informatiemodellen en afsprakenstelsels die we tijdens de verkenning zijn tegengekomen uiteengezet.

Standaard	Beschrijving
CERIF	De belangrijkste internationale standaard voor onderzoeksinformatie is CERIF, ontwikkeld en beheerd door euroCRIS. CERIF is een conceptueel model, dat een uitgebreide set aan entiteiten, relaties en attributen definieert, maar niet uitdrukt in een uitwisselingsstandaard. OpenAIRE, Pure en de Nederlandse repositories (DIDL/MODS) conformeren hun uitwisselingsstandaard aan dit model.
DIDL/MODS ¹²	De WISH werkgroep onderhoudt een profiel gebaseerd op DIDL/MODS voor nationale uitwisseling vanuit institutionele repositories, gebaseerd op CERIF. Voor praktijkgericht onderzoek zijn hierop extensies gedefinieerd, die in gebruik zijn bij o.a. SURFsharekit en HBO Kennisbank.
OpenAIRE guidelines ¹³	OpenAIRE heeft meerdere richtlijnen uitgebracht die helpen om repositories en RIS-systemen te integreren met OpenAIRE. De OpenAIRE guidelines for CRIS managers is gebaseerd op CERIF, de OpenAIRE guidelines for data archives is gebaseerd op het DataCite Metadata Schema.
Persistent identifiers	Persistent identifiers zijn een kerncomponent van de onderzoeksinfrastructuur en bieden een universele en interoperabele methode om entiteiten zoals onderzoekers, organisaties, output en projecten te identificeren, te beschrijven, en met elkaar te verbinden, en daarmee een potentieel belangrijke bron voor onderzoeksinformatie. Veel gebruikte PIDs zijn ORCID, RoR, RAiD en de DOIs van CrossRef en DataCite.
Pure Schema	Door hoge adoptie van Pure bij de universiteiten en enkele hogescholen in Nederland is het CERIF-gebaseerde metadata

¹² https://www.edustandaard.nl/standaard_afspraken/metadata-object-description-scheme-mods/metadata-object-description-scheme-versie-1-3-copy/

¹³ <https://guidelines.openaire.eu/en/latest/>

	<p>schema ook een relevante standaard. Dit schema kent echter risico's met betrekking tot de openheid en het eigenaarschap/beheer van het schema (zie POSI principles¹⁴ en de Seven Guiding Principles for Open Research Information¹⁵)</p>
SKG-IF	<p>In RDA verband wordt de Scientific Knowledge Graph Interoperability Framework (SKG-IF) ontwikkeld. Deze groep richt zich op het ontwikkelen van een framework om uitwisseling van informatie over onderzoek mogelijk te maken tussen verschillende initiatieven met betrekking tot Scientific Knowledge Graphs (o.a. repositories, databases, catalogi, LOD-collecties). Dit is de gemeenschappelijke kern van onderzoeksinformatie uit deze graphs, en zeer beperkt qua ondersteunde entiteiten, relaties en attributen.</p>

Tabel 1 Veelgebruikte informatiemodellen en afsprakenstelsels in de nationale en Europese context

Interoperabiliteit

De complexiteit van het landschap, strekkend van lokaal tot internationaal, over verschillende onderzoeksdomeinen en over verschillende toepassingsgebieden zoals PIDs, repositories en onderzoeksinformatiesystemen, maakt het uitdagend of zelfs onhaalbaar om ons volledig aan één standaard te conformeren. We zullen onderling, en naar de buitenwereld genoodzaakt zijn om op andere manieren te interopereren. In EOSC verband wordt bijvoorbeeld gewerkt om datatypes en dataschema's te identificeren, en crosswalks hiertussen mogelijk te maken. SURF is ook betrokken bij een case study, om deze technieken toe te passen voor interoperabiliteit tussen nationale PID Graphs.

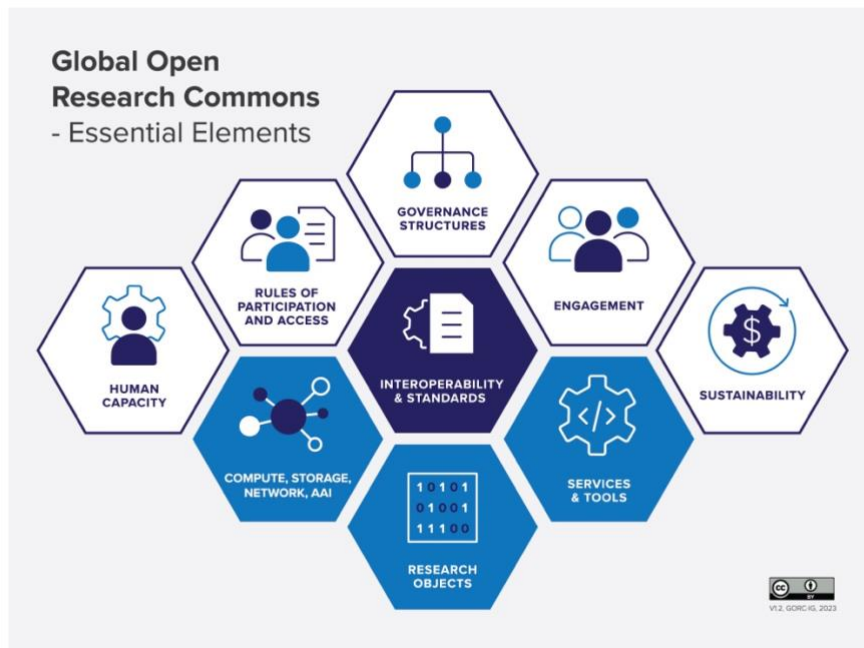
Commons, of gemeenschapsgoederen

Het ontwikkelen, beheren en implementeren van het DIDL/MODS afsprakenstelsel heeft laten zien dat dit een complex en continu proces is, met veel aspecten en stakeholders. Naast het ontwikkelen van de standaard zelf, is veel aandacht nodig voor het betrekken van de diverse stakeholders, voor besluitvorming, voor het aantrekken en ontwikkelen van relevante kennis, en voor duurzaam beheer van de standaard. Het vraagt om *collective action* en een ecosysteem aanpak.

¹⁴ <https://openscholarlyinfrastructure.org/>

¹⁵ Bijsterbosch, M. et al. (2022). *Seven Guiding Principles for Open Research Information* (Version 1). <https://doi.org/10.5281/zenodo.6074944>

De Global Open Research Commons (GORC)¹⁶ zijn ontwikkeld als een taal en model voor 'research commons', vergelijkbaar met EOSC. SURF ziet dit model als toepasbaar voor soortgelijke initiatieven en experimenteert ermee. Daarnaast wordt het beschouwd als een geschikt model om de multidisciplinariteit en *collective action* rondom onderzoeksinformatie te adresseren.



Figuur 2: Typologie, bestaande uit 9 essentiële elementen, voor een 'Global Open Research Commons'.

¹⁶ Payne, K. et al. (2023). *The Global Open Research Commons International Model Report, Version 1 (1.0)*. <https://doi.org/10.15497/RDA00097>

3. Use case: Publinova

We gebruiken Publinova als eerste use case om -iteratief- het domein van onderzoeksinformatie in het hbo te adresseren. Publinova ondersteunt een select aantal entiteiten, relaties en attributen en is daarmee een goede eerste use case om de benodigde standaarden te verkennen en te onderzoeken hoe die toegepast kunnen worden. Hierbij zijn we ons steeds bewust dat deze use case een beperkte scope heeft, en niet representatief is voor de gehele informatiebehoefte rondom onderzoeksinformatie in praktijkgericht onderzoek. We zullen dan ook bespreken wat dit betekent voor onze resultaten en bevindingen.

Wat is Publinova?

Publinova is het nationale platform waar praktijkgericht onderzoek zichtbaar en vindbaar wordt. Het platform is ontstaan uit een project in opdracht van 14 hogescholen met als doel om de kennisinfrastructuur voor praktijkgericht onderzoek te verbeteren en aan te sluiten bij de Open Science ambities. Voor het realiseren van Publinova is samengewerkt met de HBO Kennisinfrastructuur, de Vereniging Hogescholen, Regieorgaan SIA en SURF. Publinova is inmiddels een dienst van SURF en wordt samen met SIA en de hogescholen doorontwikkeld.

In 2023 is het platform live gegaan en sindsdien hebben hogescholen de mogelijkheid om informatie over onderzoekers, onderzoeksprojecten en resultaten te presenteren. Op dit moment zijn 13 hogescholen aangesloten op het platform die data aanleveren vanuit hun eigen bronsystemen. Hier ligt ook meteen een belangrijke uitdaging: ieder bronsysteem heeft een eigen datamodel en informatie wordt dus niet op een gestandaardiseerde manier aangeleverd. Op lange termijn is het niet beheersbaar als elk bronsysteem en elk afnemend systeem ad hoc integraties moet onderhouden. In dit hoofdstuk laten we zien hoe we de uitwisseling van onderzoeksinformatie naar Publinova kunnen standaardiseren.

Informatiebehoefte en waardeketen

Om de uitwisseling naar Publinova te verbeteren kijken we eerst naar de waardeketen: Het profileren van praktijkgericht onderzoek, ten bate van het hergebruik van haar kennis en producten. We zien dat de informatie die Publinova verzamelt en presenteert invulling geven aan deze waarde, we laten echter buiten beschouwing of informatie ontbreekt. In Appendix I – IV koppelen we de informatie in Publinova aan deze waardeketen (zie laatste kolom).

De profilering van praktijkgericht onderzoek vraagt andere informatie dan de huidige HBO Kennisbank of een voormalig portal als NARCIS: de doelgroep is iets anders, en daarmee de wijze waarop informatie wordt ge(re)presenteerd. Dit is te zien aan bijvoorbeeld het tonen van contactgegevens en mogelijkheden om met titels en beschrijvingen expertise te profileren. Waar bij NARCIS en HBO Kennisbank het

product/resultaat centraal staat, kan een onderzoeker zich op Publinova onafhankelijk van producten of resultaten profileren.

Informatiemodel en afsprakenstelsel

Bij de ontwikkeling van Publinova lag de focus initieel op het aantonen van de waarde voor gebruikers en instellingen, en niet op het ontwikkelen van een uitwisselingsstandaard. Informatie wordt deels via DIDL/MODS, maar ook op andere manieren zoals web-scraping opgehaald. Het is voor Publinova en voor instellingen van belang dat er een uitwisselingsstandaard met bijbehorende afspraken wordt ontwikkeld, zodat zij hier rekening mee kunnen houden bij het inrichten van hun toekomstige processen en systemen.

Voor deze use case zijn we uitgegaan van het interne datamodel van Publinova. We startten met het vergelijken van de persoon met de in ontwikkeling zijnde nieuwe versie van CERIF, en breidden dit uit naar andere entiteiten (partij, product en project), en met andere standaarden (DIDL/MODS, SKG-IF). Zie Appendix I - IV voor de mapping die we gemaakt hebben van de velden in Publinova op CERIF per entiteit. Dit is de minimale set aan onderzoeksinformatie voor de use case Publinova, en deze zal verder doorontwikkeld moeten worden voor andere use cases.

De volgende zaken vallen op:

- *Dekking.* Veel velden van Publinova komen overeen met bijbehorende entiteiten en attributen in CERIF. Dit is te verklaren doordat veel velden common practice zijn, en ook reeds worden toegepast in o.a. de bestaande DIDL/MODS afspraken.
- *Affiliaties.* In Publinova wordt de organisatie afgeleid van de aanleverende instelling, en is functie een enkelvoudig veld. In CERIF en andere standaarden zie je de mogelijkheid om de relatie met een organisatie, incl. functie per organisatie, in één of meerdere affiliaties weer te geven, met begin- en einddatum wanneer er meerdere of opeenvolgende functies zijn.
- *Beschrijving en onderzoeksthema's.* In CERIF zijn hiervoor nog geen passende attributen. We verklaren dit door de relatief nieuwe toepassing van profileren.
- *Privacy.* Voor het profileren van de onderzoekers wordt gebruik gemaakt van een profielfoto, mailadres, en telefoonnummer. Op dit moment worden foto's niet ondersteund in CERIF.

We hebben ook verkend wat voor afspraken er nodig zijn, enerzijds vanuit de bestaande DIDL/MODS afspraken, maar ook door te kijken naar de content van Publinova.

- *Semantiek.* Welke rollen worden er gebruikt voor onderzoekers? Kan dit zo simpel als lector en onderzoeker, of moet dit uitgebreider? Datzelfde geldt voor andere relaties tussen de entiteiten (zoals de rol die iemand had bij

totstandkoming van een product, de rol in een project, etc.). Hier zijn de waardeketens en use cases nodig om te bepalen welke rollen nodig zijn.

- *Producttypen*. Publinova maakt op dit moment gebruik van Mime-type voor het classificeren van producttypen. Dit is voor vindbaarheid en hergebruik geen logische keuze. Er moeten afspraken worden gemaakt over de set aan producttypen, we kunnen beginnen met de producttypen zoals beschreven in de Semantics¹⁷ afspraken.
- *Persistent Identifiers*. Welke identifiers gebruiken we voor unieke identificatie van personen, projecten, organisaties en producten? Bij elk van de entiteiten zijn er uitdagingen, die hieronder uiteengezet worden.
- *Lifecycle of status van entiteiten*. Er moeten afspraken worden gemaakt over wijzigingen in bijv. dienstverband, het wijzigen/opheffen van afdelingen of de verschillende stadia van projecten. Welke partners worden opgenomen, en wat wordt daarvan weergegeven?
- *Privacy*. Omdat profileren en hergebruik een grote rol speelt in Publinova, wordt het benaderen van onderzoekers ondersteund. Echter vraagt dit om duidelijke afspraken over waar en wanneer contactgegevens gedeeld mogen worden.

Persistent identifiers

Eén van de lastigste taken van een aggregator als Publinova is het ontdebellen van informatie afkomstig uit de verschillende bronsystemen. Hoe kun je identificeren of twee records naar hetzelfde document verwijzen als er geen unieke identifier is toegekend? Een mogelijkheid is om te matchen op basis van titel, maar dit gaat snel fout als de titel net op een andere manier is ingevoerd (bijvoorbeeld met of zonder ondertitel). Om deze reden is het gebruik van persistent identifiers (PIDs) noodzakelijk voor het maken van een duurzame/uniëke verwijzing naar entiteiten.

In Tabel 2 hieronder geven we voor elk van de entiteiten in Publinova (persoon, project, partij, product) aan welke PID wij adviseren te gebruiken, maar ook wat de uitdagingen zijn.

Entiteit	PID	Uitdaging
Persoon	ORCID ¹⁸	Het meestgebruikte PID systeem voor (actieve) onderzoekers is op dit moment ORCID. Overige zijn o.a. ISNI, DAI, Scopus ID en ResearcherID. DAI is in Nederland nog in gebruik, maar wordt als standaard niet meer onderhouden. Een ORCID moet worden aangemaakt door de onderzoekers zelf en de adoptie van ORCID in het praktijkgericht onderzoek is nog niet hoog.

¹⁷ https://www.edustandaard.nl/standaard_afspraken/semantics/semantics-versie-1-0/

¹⁸ <https://info.orcid.org/researchers/>

Project	RAiD ¹⁹	Research Activity Identifier (RAiD), een persistent identifier voor projecten, is in ontwikkeling (door ARDC) en SURF gaat het 'European Registration Agency' hosten. In 2024 zullen meerdere pilots starten (ook binnen het praktijkgericht onderzoek) waarin het gebruik van RAiD wordt getest.
Partij	ROR ²⁰	ROR is een internationaal geaccepteerde unieke identifier voor onderzoeksinstituten die een juridische status hebben. Een nadeel is echter dat onderzoekseenheden zoals Centres of Expertise of lectoraten geen ROR kunnen krijgen. Hier moet een oplossing voor worden gevonden.
Product	DOI	De Digital Object Identifier wordt gebruikt om producten uit onderzoek uniek te identificeren. CrossRef registreert DOIs voor artikelen in wetenschappelijke tijdschriften, DataCite doet dit voor datasets. Ook repositories zoals SURFsharekit of Zenodo bieden de mogelijkheid om DOIs toe te kennen aan producten. Artikelen in vaktijdschriften hebben echter vaak geen DOI. Hetzelfde geldt voor grijze literatuur zoals beleidsstukken en technische rapporten, terwijl dit wel veelvoorkomende output is in praktijkgericht onderzoek.

Tabel 2 Lijst van aanbevolen PIDs voor de vier entiteiten in Publinova, met bijbehorende uitdagingen.

Commons

In de werkgroep werd snel duidelijk dat er veel komt kijken bij het ontwikkelen, beheren en implementeren van een uitwisselingsstandaard. Zo werd ons ook duidelijk in gesprek met de trekkers van de DIDL/MODS afspraken. We hanteren de 9 *Essential Elements* van de GORC typologie als framework om de opgedane inzichten te duiden:

- *Human Capacity*. De werkgroep ervaart dat de combinatie van beschikbare tijd en verschillende soorten kennis (technisch, interoperabiliteit, architectuur, toepassing) nog onvoldoende is om sturing en uitvoering te geven aan goede uitwisseling van onderzoeksinformatie in praktijkgericht onderzoek. Het is van belang dat we deze kennis blijven opbouwen, uitwisselen en toepassen, maar ook dat we capaciteit delen met initiatieven in wetenschappelijk onderzoek.
- *Governance structure*. Met de WISH is er op dit moment een governance structuur voor de DIDL/MODS afspraken. Deze sluit niet volledig aan op de behoefte vanuit praktijkgericht onderzoek voor bijv. niet-tekstuele publicaties, of

¹⁹ <https://faircore4eosc.eu/eosc-core-components/research-activity-identifier-service-raid>

²⁰ <https://ror.org/>

voor andere entiteiten dan output. Hier loont om te onderzoeken hoe de WISH kan worden hergebruikt en uitgebreid, dan wel hoe hiermee kan worden samengewerkt.

- *Engagement.* Het is belangrijk om alle stakeholders te betrekken, van instellingen en diensten die de informatie (willen) aanleveren, tot instellingen en diensten die de informatie gebruiken. Vanuit de werkgroep is samengewerkt met Publinova, SURFsharekit en HBO Kennisbank. Via euroCRIS en EOSC zoeken we best practices en afstemming met vergelijkbare Europese initiatieven. Daarnaast wordt gezocht naar aansluiting bij de behoeften en oplossingen binnen wetenschappelijk onderzoek.
- *Rules of Access and Participation.* Er zullen afspraken gemaakt moeten worden over wie gebruik mag maken van deze onderzoeksinformatie (o.a. regels voor toegang en privacy).
- *Interoperability & Standards.* Het is niet mogelijk om alle interoperabiliteitsvraagstukken op te lossen met één standaard en er zullen meerdere standaarden naast elkaar bestaan.
- *Sustainability.* Om de standaarden en afspraken te ontwikkelen, implementeren en beheren moet er op een duurzame manier in geïnvesteerd worden.
- *Research Objects.* Er is een duidelijke ontwikkeling van een uitwisselingsstandaard voor publicaties (output) naar de vier entiteiten en hun relaties.
- *Services and Tools.* De uitwisselingsstandaard zal sterk verweven zijn met de repositories en de (nog te implementeren) RIS systemen bij de instellingen.
- *Network, Compute, Storage.* Deze elementen zijn bij deze case study niet of minimaal ter sprake gekomen.

4. Bevindingen en aanbevelingen

Dit adviesrapport doet op basis van de verkenning en use case aanbevelingen aan de hbo-instellingen, koepelorganisaties, SURF, en de community. Door nationale en internationale ontwikkelingen te volgen en in te gaan op de use case Publinova komen we tot de conclusie dat de huidige standaarden en afspraken niet toereikend zijn voor de informatiebehoefte in het hbo.

In de korte secties hieronder worden de bevindingen en aanbevelingen uit de verkenning en use case kort samengevat:

Bevinding 1

Het opstellen, beheren en implementeren van de standaard vraagt om actieve participatie van de verschillende stakeholders, met bijbehorende capaciteit (kwantitatief en kwalitatief).

Bevinding 2

We zien geen fundamentele verschillen met onderzoeksinformatie in het wo. De andere producttypen zijn wel kenmerkend, maar niet uniek voor praktijkgericht onderzoek. Dat geldt ook voor bijvoorbeeld de attributen die Publinova gebruikt om te profileren, zoals de beschrijving van de expertise of contactgegevens van een persoon.

Bevinding 3

Het uitgevoerde werk heeft betrekking op een enkele use case, het profileren en hergebruik via Publinova. We schatten in dat de inzichten ervan breder toepasbaar zijn. Voor andere use cases, met name die binnen de waardeketen “Sturen en verantwoorden”, verwachten we additionele informatiebehoeften, die zullen vragen om uitbreiding van de entiteiten en attributen, en het bijbehorende afsprakenstelsel.

Bevinding 4

Internationaal is CERIF de de facto standaard, maar deze is alleen conceptueel. Er is een concrete implementatie nodig en daarvoor is geen standaard beschikbaar waar we ons volledig aan kunnen conformeren. Voor tekstuele output is DIDL/MODS binnen Nederland een bruikbare standaard, maar deze is wel aan vernieuwing toe. Voor overige entiteiten zijn relevante initiatieven in bijvoorbeeld Finland en België.

Bevinding 5

Persistent identifiers zijn een waardevol instrument om producten, personen, projecten en partijen te onderscheiden, beschrijven, koppelen en uit te wisselen. Hiermee kan de kwaliteit van onderzoeksinformatie worden vergroot en de administratieve last worden verlaagd.

Bevinding 6

De vocabulaires, bijvoorbeeld die voor producttypen en onderzoeksgebieden, moeten worden bijgesteld om aan te sluiten bij de veranderende behoefte van de community en standaarden in het veld.

Aanbevelingen voor het DCC-PO

Organiseer eigenaarschap en opdracht, met daarin:

- Concrete use cases, ook voor “Sturen en verantwoorden” (aan welke sturingsinformatie is urgente behoefte?)
- Ambitie om zoveel mogelijk op te trekken met wo als het gaat om uitwisselingsstandaarden, voor efficiënte inzet van beschikbare capaciteit en maximale interoperabiliteit.
- Ambitie om te werken met open standaarden, zodat we onafhankelijk van applicaties beslissingen kunnen maken over eigen gegevens.

Aanbevelingen voor de instellingen

Zorg voor:

- Commitment voor benodigde capaciteit van de verschillende stakeholders binnen de instelling (bestuurlijk, beleidsmatig, technisch, en operationeel).
- Samenwerking tussen de instellingen, o.a. via het DCC-PO.

Aanbevelingen voor de community

- Maak standaarden of profielen interoperabel door deze op CERIF te baseren, en door gebruik te maken van persistent identifiers.
- Gebruik bestaande standaarden zoals DIDL/MODS en de Semantics afspraken als opstap en ontwikkel ze door.
- Onderken dat er meerdere standaarden of profielen nodig kunnen zijn voor verschillende toepassingen.
- Bepaal welke vocabulaires nodig zijn voor de informatiebehoefte van de gebruikers van onderzoeksinformatie (zoals voor producttypen, rollen binnen projecten en producten, en voor onderzoeksgebieden/disciplines) en bouw hierbij voort op bestaande vocabulaires.

Aanbevelingen voor SURF

Vervul de regierol, met de volgende aandachtspunten:

- Faciliteer de verbinding met wo, o.a. door praktijkgericht onderzoek en wetenschappelijk onderzoek aan de universiteiten samen te betrekken in initiatieven (zoals reeds gebeurt bij RAiD pilots en Machine Actionable DMP).
- Leg verbinding met internationale ontwikkelingen (zoals reeds gebeurt in EOSC verband).
- Betrek diensten als SURFsharekit, Publinova, en HBO Kennisbank actief bij het opstellen, testen en implementeren van de (tussen)resultaten van de standaard.
- Stuur op de ontwikkeling en borging van een zelfstandige community met kennis, doorontwikkeling, en implementatie van de standaard in een nationale infrastructuur.

Bijlage 1 Tabel Entiteit Persoon

Attribuut	Definitie	Publinova veld	CERIF	Waarde
Naam	De volledige naam van de persoon	Full_name: "string"	Person_Name	Vinden, Herkennen
Voornaam	De officiële voornaam van de persoon	first_name: "string",	First_Names	Vinden, Herkennen
Achternaam	De officiële achternaam van de persoon	last_name: "string",	Family_Names	Vinden, Herkennen
Functie	De functie die de persoon binnen de organisatie vervult	job_title: "string"	Affiliation_Statement	Vinden, Herkennen, Beoordelen
Gerelateerde organisaties	De organisatie-eenheden waarmee deze persoon geaffilieerd is	parties: ["string", ..., "string"]	Affiliation_Statement	Vinden, Herkennen, Beoordelen
Beschrijving	Een beschrijving van het expertisedomein en specialiteiten van deze persoon	description: "string"	n/a	Vinden, Beoordelen
Expertise	Beschrijvende termen (keywords) die de onderzoeksactiviteiten van de persoon uitdrukken	skills: ["string", ..., "string"]	Expertise_and_Skills	Vinden, Beoordelen
Identificer	Unieke externe identificer van de persoon	orcid: "string"; dai: "string"; isni: "string"	ORCID_iD; ISNI_iD;	Vinden, Herkennen
Emailadres	Het emailadres waarop de persoon bereikt kan worden (email, telefoon)	Email: "string"	Electronic_Address > Email_Address	Benaderen
Telefoonnummer	Het telefoonnummer waarop de persoon bereikt kan worden	Phone: "string"	Electronic_Address > Telephone_Number	Benaderen
Onderzoeksthema's	Set van onderzoeksthema's die gebruikt worden om onderzoek te classificeren (ook gebruikt in HBO kennisbank)	research_themes: ["string", ..., "string"]	n/a	Vinden

Bijlage 2 Tabel Entiteit Project

Attribuut	Definitie	Publinova veld	CERIF	Waarde
Titel	De titel van het project	"title": "string",	Title	Vinden, Herkennen
Beschrijving	Een beschrijving van het project	"description": "string",	cfProjAbstr (1.6) ²¹	Vinden, Herkennen, Beoordelen
Doel	Een korte beschrijving van het doel van het project	"goal": "string",	n/a	Vinden, Herkennen, Beoordelen
Trefwoorden	Beschrijvende termen (keywords) die de onderzoeksactiviteiten van het project uitdrukken	"keywords":["string",...,"string"],	cfProjKeyw (1.6)	Vinden
Onderzoeksthema's	Set van onderzoeksthema's die gebruikt worden om onderzoek te classificeren (ook gebruikt in HBO kennisbank)	"research_themes":["string",...,"string"]	n/a	Vinden
Startdatum	De startdatum van het project	started_at: "string",	cfStartDate (1.6)	Vinden, herkennen
Einddatum	De einddatum van het project	ended_at: "string",	cfEndDate (1.6)	Vinden, herkennen
Status	De status van het project op basis van een vaste lijst	status: "string",	n/a	Vinden, herkennen
Foto	Een foto gebruikt voor het profileren van het project	photo_url: "string",	n/a	Herkennen
Identificer	Unieke en persistente identificer van het project	"external_id": "string"	cfProjId (1.6)	Vinden

²¹ We hebben de meest recente "Refactored" versie van CERIF gebruikt. Deze versie is echter in ontwikkeling, dus waar nodig zijn we uitgeweken naar de volledig gedocumenteerde 1.6 versie.

Bijlage 3 Tabel Entiteit Partij

Attribuut	Definitie	Publinova veld	CERIF	Waarde
Naam	De naam van de organisatie-eenheid	Eigen invoer	Organisation_Unit > Name	Herkennen
Acroniem	Het acroniem van de organisatie	Eigen invoer	Organisation_Unit > Acronym	Herkennen
Beschrijving	Een beschrijving van de onderzoeksactiviteiten van deze organisatie	Eigen invoer	Organisation_Unit > Description	Herkennen, Beoordelen
Type	Het type organisatie gebaseerd op een vaste lijst.	type: "string",	cfOrgUnit_Class (1.6)	Vinden, Herkennen
Logo	Het logo van de organisatie-eenheid	photo_url: "string",	n/a	Herkennen
Identificer	Unieke en persistente identificer van de organisatie-eenheid	external_id: "string",	ROR_Identificer	Vinden

Bijlage 4 Tabel Entiteit Product

Attribuut	Definitie	Publinova veld	CERIF	Waarde
Titel	De titel van het product	"title": "string",	Title	Vinden, Herkennen
Taal	De taal van het product	"language":	Language	Vinden, Herkennen
Identifiers	Een unieke externe identifier van het product	external_id: "string",	DOI_Identifier	Vinden
Publicatiedatum	De datum waarop het product gepubliceerd is	"published_at": "string",	cfResPublDate	Vinden, Beoordelen
Auteur	De auteur(s) van het product.	"authors": [{persoon object}, ... {persoon object}]	cfPers_ResProd > cfClass-Pers_ResProd (1.6)	Vinden, Herkennen, Beoordelen, Benaderen
Eigenaar	De person(en) met beheerrechten van het product	"owners": [{persoon object}, ... {persoon object}]	cfPers_ResProd > cfClass-Pers_ResProd (1.6)	Vinden, Herkennen, Beoordelen, Benaderen
Contactpersoon	De contactperson(en) voor het product	"contacts": [{persoon object}, ... {persoon object}]	cfPers_ResProd > cfClass-Pers_ResProd (1.6)	Benaderen
Gerelateerde partijen	De partijen waar het product aan gekoppeld is.	parties: ["string", ..., "string"]	cfOrgUnit_ResProd (1.6)	Herkennen
Beschrijving		"description": "string",	cfResProdDescr (1.6)	Herkennen, Beoordelen
Trefwoorden	Beschrijvende termen (keywords) die de onderzoeksactiviteiten van het product uitdrukken	"keywords":["string",..., "string"],	cfResProdKeyw (1.6)	Vinden, Herkennen
Onderzoeksthema's	Set van onderzoeksthema's die gebruikt worden om onderzoek te classificeren (ook gebruikt in HBO)	"research_themes": ["string",..., "string"]	cfResProdClass (1.6)	Vinden

	kennisbank)			
Type	Het soort product op basis van Mime-type.	type: "document",	cfResProf_Class (1.6)	Vinden
Copyright license	De copyright license van het product	"copyright": "string",	cfDCRightsMMLicens e (1.6)	Hergebruiken