

Open Research Information Agenda

(voorheen Business Case Open Knowledge Base)

Versie 1.0 d.d. 11-11-2022

Managementsamenvatting

Om digitale soevereiniteit en kwaliteit te borgen op rapportage, evaluatie en vindbaarheid van Nederlands onderzoek is op verzoek van UNL een traject gestart waarin een business case is opgesteld voor een 'Open Knowledge Base' voor informatie over onderzoek. Uit dit traject, gecoördineerd door SURF en UNL is gebleken dat de behoefte aan open onderzoeksinformatie breed wordt gedragen. Dit is bevestigd in een workshop met een brede afvaardiging van nationale belanghebbenden in juni 2022. Uit de studie is gebleken dat er consensus over de open benadering bestaat en dat via diverse projecten iteratief toegegroeid moet worden naar een open onderzoeksinformatie agenda. Daarbij is een iteratieve benadering ook noodzakelijk om nader te bepalen of er meerwaarde is voor een centrale 'knowledge base of knowledge graph' dan wel centrale diensten en hoe verbeteringen van lokale en internationale bronnen bewerkstelligt kan worden, gebruikmakend van zoveel mogelijk open databronnen.

Ten eerste zijn er afspraken nodig die garanderen dat we in Nederland (en internationaal) dezelfde definities (denk bijvoorbeeld aan Open Access) en standaarden (denk aan het gebruik van Persistent Identifiers) hanteren zodat informatie ook eenduidig geregistreerd wordt. Alleen dan is het aggregeren van onderzoeksinformatie over instellingsgrenzen heen mogelijk en kan deze informatie zinvol worden gebruikt voor de eigen en nationale inzichten, rapportage en evaluatiedoeleinden en voor de verbetering van gebruikte (open) databronnen. Bovendien moeten we deze afspraken ook bij de marktpartijen kunnen afdwingen waar dat nog niet het geval is. Ten tweede zijn niet alle brondata open en momenteel nog van onvoldoende kwaliteit en volledigheid om alle gewenste toepassingen van open onderzoeksinformatie te realiseren. Een kwaliteitsimpuls is hier nodig om te zorgen dat én de instellingssystemen én andere (inter)nationale bronnen ook kwalitatief goede informatie kunnen bieden.

Om ook te verkennen welke toepassingen we met bestaande lokale bronnen en beschikbare open onderzoeksinformatie al kunnen faciliteren en welke niet, zijn meerdere (pilot)projecten nodig. Deze zullen iteratief bijdragen aan een groeiende Open Research Information Agenda en tegelijkertijd de kwaliteit van instellingssystemen en andere open bronnen verbeteren. Daarnaast zullen projecten gestart moeten worden die bestaande overzichten van onderzoeksinformatie verbeteren dan wel uitbreiden (denk bv aan uitbreiding open access monitor naar conferentiebijdragen en boeken).

Om deze reden wordt vanuit de studie geconcludeerd dat er gezamenlijk/nationaal naar een Open Research Information Agenda (ORIA) toegewerkt moet worden in het kader van de gewenste/beoogde transitie en dat daartoe gezamenlijke investeringen nodig zijn, zowel lokaal als (inter)nationaal. Daarbij is het voorstel om dit eerst in 2023 verder te verkennen via een projectorganisatie.

1. Waarom van open research information agenda

1.1 Over welke type onderzoeksinformatie hebben we het

Het belang van betrouwbare en herbruikbare onderzoeksinformatie is groeiende en commerciële partijen bieden in toenemende mate services aan rond informatie over wetenschappelijk onderzoek uitgevoerd door publieke kennisinstellingen. Het gaat in deze over informatie over publiek gefinancierd onderzoek in Nederland d.w.z. beschrijvende informatie en bijhorende (impact)evaluaties over onderzoeksprojecten (projectbeschrijvingen, subsidies, enz), onderzoeksresultaten (publicaties, datasets, onderzoekssoftware en leermaterialen) en betrokken academici.

Het betreft onder andere informatie en analyses over onderzoek ten behoeve van:

- A. managementinformatie/rapportages om zowel op niveaus van teams, instellingen, netwerken, consortia, etc. dan wel landelijk niveau over onderzoeksoutput, de impact van wetenschap dan wel rapportage over afgesproken KPI's¹ (bijv. monitoring open access) (met subsidiegever, overheid);
- B. indicatoren voor impact van (wetenschappelijk) onderzoek zoals publicaties, datasets maar ook co-creatie, en deelname aan publieke discussie;
- C. geaggregeerde informatie die gebruikt kan worden voor evaluatie en beoordelingen van onderzoeksaanvragen, van resultaten onderzoeksprojecten, van onderzoeksgroepen (bv visitaties, SEP²);
- D. het efficiënter aanleveren van metadata ten aanzien van onderzoeksprojecten door onderzoekers aan subsidiegevers, aan uitgevers waar ze publiceren en aan de eigen werkgever(s);
- E. het genereren van formatieve en summatieve evaluaties waarbij onderzoeksgroepen en instellingen nationaal met elkaar vergeleken kunnen worden;
- F. communicatie over onderzoek met het brede publiek (bedrijven, citizen scientists, politici, journalisten, etc.); het vindbaar en herbruikbaar maken van onderzoeksoutput tot voor het brede publiek aantonen wat de (maatschappelijke) impact is, en hoe het wetenschappelijk systeem werkt en de impact tot stand komt;
- G. profielpagina's voor onderzoekers, onderzoeksgroepen, onderzoeksteams, ter profilering van hun prestaties en het vergroten van vindbaarheid als onderzoeker voor (nieuwe) samenwerkingsverbanden en voor hun verdere loopbaan mogelijkheden.

1.2 Belang van open beschikbaarheid onderzoeksinformatie

Op Europees niveau is men het over het algemeen eens dat resultaten van wetenschappelijk onderzoek zoveel mogelijk open beschikbaar moeten zijn. Het is dan ook een logisch gevolg dat ook de informatie over het onderzoek zo compleet en zo open als mogelijk beschikbaar gesteld moet worden ten behoeve van het hergebruik door alle belanghebbenden. Ook internationaal zien we de beweging dat onderzoeksinformatie open moet zijn (zie o.a. [ICOLC Statement on the Metadata Rights of Libraries](#)) en waarbij verschillende landen dit centraal dan wel nationaal oppakken (zie [Research Information Finland](#), [Research portal Denmark/NORA](#), [HAL France](#), [FRIS research portal](#)). In Nederland speelt Centre for

¹ KPI = kritische prestatie-indicator

² SEP = Strategy Evaluation Protocol

Science and Technology Studies (CWTS) als expertisecentrum een leidende rol op gebied van onderzoeksinformatie.

Het zoveel mogelijk open beschikbaar zijn van data en onderzoeksinformatie is algemeen beleid en is ook uitgangspunt van de Werkagenda "Waardengedreven Digitaliseren" in Nederland³. Dit houdt in dat we als Nederlandse kennispartijen een actieve agenda moeten voeren over open onderzoeksinformatie. Hierbij past de statement "as open as possible and as closed as necessary" die afkomstig is uit de beweging rond FAIR data. Hoewel de [FAIR data principles](#) met name zijn toegespitst op het herbruikbaar maken van onderzoeksdata (gegevens *uit* onderzoek) en niet specifiek op onderzoeksinformatie (informatie *over* onderzoek), gaan deze principes voor een groot deel juist over metadata afspraken. En juist op het gebied van metadata afspraken raken onderzoeksdata en onderzoeksinformatie elkaar. Denk daarbij aan het gebruik Persistent Identifiers, relaties tussen verschillende informatie entiteiten en provenance informatie.

1.3 Belang van volledigheid en voldoende kwaliteit

De benodigde informatie is nu vaak niet volledig en/of soms van onvoldoende kwaliteit (vaak dubbelingen bij aggregatie over instellingen heen). Kwaliteit is zeer nodig omdat rapportages betrouwbaar moeten zijn, informatie gebruikt wordt bij beoordelingen van subsidies en onderzoeksgroepen (SEP) en ook bij financiële transacties.

1.4 Kwantitatieve rapportages binnen kwalitatieve narratives

Betrouwbare, valideerbare en herbruikbare informatie over onderzoek is voor publieke kennisinstellingen, overheden en andere publieke partijen essentieel (zoals in paragraaf 1.1 besproken in punten A-G) om output en impact te demonstreren. Het belang van betrouwbare en herbruikbare onderzoeksinformatie zal in de toekomst niet minder worden door toename in diversiteit aan methodes om impact te demonstreren (zie bijvoorbeeld de beweging die de [DORA declaration](#) in gang heeft gezet).

Daarbij hebben de kennisinstellingen aangegeven dat onderzoeksevaluatie en analyses niet louter op basis van kwantitatieve indicatoren kunnen worden uitgevoerd. In deze ontwikkeling is onder andere ingezet op verschuiving naar een andere wijze van [research assessment](#), met meer nadruk op kwalitatieve gegevens en narratives. Desalniettemin wordt ook onderkend dat (andersoortige/verrijkte) kwantitatieve indicatoren van groot belang blijven. Voor zowel de kwantitatieve als kwalitatieve analyses is echter breed beschikbare, betrouwbare en herbruikbare onderzoeksinformatie van groot belang. Een logisch uitgangspunt is dus dat dit soort informatie (zoals in paragraaf 1.1 besproken in punten A-G) open beschikbaar moet zijn in het kader van Open Research Information Agenda (ORIA).

1.5 Probleem van vendor lock-in bij beperkt aantal partijen

Bovendien zijn instellingen voor onderzoeksinformatie afhankelijker geworden van een beperkt aantal commerciële partijen. Het gevolg is dat de digitale en academische soevereiniteit onder druk komt te staan. Deze zorgen zijn ook al in 2019 geuit door de rectoren van de Nederlandse universiteiten. Begin

³ Werkagenda "Waardengedreven Digitaliseren" (https://www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2022Z21101&did=2022D45419) stelt o.a. dat "waar dat kan werken in open source, met open data, open algoritmen en open beleidsregels".

2020 zijn deze zorgen door UNL geadresseerd en is een brede nationale [Taskforce](#) ingesteld om deze zorgen en de kwestie van openheid van onderzoeksinformatie te adresseren.

1.6 Twee sporen om uitdagingen en zorgen aan te pakken

Deze Taskforce heeft een tweesporenbeleid [geadviseerd](#) dat door alle partijen verbonden aan het Nationaal Programma Open Science (NPOS) begin 2021 is omarmd:

1. De voorwaarden en eisen vanuit publieke kennisinstellingen vertaald naar ‘Seven Guiding Principles’ voor open onderzoeksinformatie welke de academische gemeenschap van toepassing wil laten zijn voor alle publieke en private services. Deze zijn als eerste toegepast binnen het 2020-2024 contract met Elsevier. Na een open consultatie zijn deze Guiding Principles doorontwikkeld ([V2.1. van begin 2022](#)) en worden nu ook op Europees niveau opgepakt (o.a. [LERU data statement](#), onderdeel nieuwe ERA actielijnen).
2. Het advies om te investeren in de meeste kritieke onderzoeksinformatie; een publieke infrastructuur die voor iedereen open toegankelijk is en onder publieke sturing staat. Hiertoe is in opdracht van die Taskforce [een haalbaarheidsstudie](#) uitgevoerd naar een “Open Knowledge Base” (OKB). N.a.v. deze studie hebben kennisinstellingen/verbonden aan NPOS aangegeven het belang ervan te onderschrijven, en dat e.e.a. nader uitgewerkt dient te worden in een businesscase. Dit hebben UNL en SURF gezamenlijk opgepakt gedurende de periode jan-aug 2022.

Bijlage 6 geeft een overzicht van gemaakte stappen tot nu toe.

2. Vormgeven van Open Research Information Agenda

In de periode januari-juli 2022 zijn door een werkgroep, bestaande uit medewerkers van UNL en SURF, gesprekken gevoerd met diverse (inter)nationale stakeholders (zie bijlage 1). Op basis van eerste bevindingen is op 13 juni 2022 een workshop met circa 35 experts georganiseerd (zie bijlage 2).

Op basis van die gesprekken, de workshop en voortschrijdend inzicht wordt geconcludeerd dat er landelijk geïnvesteerd moet worden in een ORIA met de volgende onderdelen:

- Afsprakenstelsel
- Kwaliteitsimpuls
- Pilots en projecten

2.1 Afsprakenstelsel

Om onderzoeksinformatie ten behoeve van A-G (zoals in paragraaf 1.1 besproken) beschikbaar te stellen dienen afspraken gemaakt te worden voor betrouwbaarheid, herbruikbaarheid en open beschikbaarheid van deze informatie. Voorbeelden waar afspraken moeten worden gemaakt of geüpdate zijn de vocabulaires en definities voor eenduidig en betrouwbare analyses. Dergelijke voorbeelden van afspraken zijn te vinden in Edustandaarden, HORA⁴/HOSA⁵, Guiding Principles, etc.

Gemeenschappelijke afspraken dienen ook gemaakt worden tussen bijvoorbeeld UNL, VH, KNAW, NWO en ZonMW over voeden en gebruik van onderzoeksinformatie uit diverse (inter)nationale open

⁴ HORA = Hoger Onderwijs Referentie Architectuur

⁵ HOSA = Hoger Onderwijs Sector Architectuur

bronnen. En ook hoe die informatie vanuit centrale pilots en projecten terugvloeit naar verbeteringen van eigen systemen (o.a. CRIS⁶) en andere bronsystemen. Bij UNL geldt dit bijvoorbeeld voor de (vernieuwde) Open Access monitor, de financiële transacties rond deals met uitgevers, jaarlijkse KUOZ⁷ uitvraag, de rapportages nodig voor het SEP en KPI-rapportage aan het ministerie van OCW.

De, al gemaakte, basisafpraak is dat onderzoeksinformatie open moeten zijn en dat we daarop de Guiding Principles toepassen, zowel naar commerciële serviceproviders als m.b.t. onze eigen systemen. Daarbij is uitgesproken dat het eigenaarschap over de metadata, informatie en analyses over publiek gefinancierd onderzoek per definitie in het publieke domein ligt. Dit is zo bevestigd met Elsevier in het lopende contract maar nog niet met andere commerciële partijen.

Ook dienen publieke kennispartijen onderling af te spreken dat zij deze informatie (of iig welke van deze) en de verrijkingen op deze informatie (zoals de mate van open access, indicatoren die iets zeggen over maatschappelijke impact) openbaar beschikbaar stellen. Aan de universiteiten, en andere kennisinstellingen, wordt daarom gevraagd zo snel mogelijk, en waar mogelijk, hun onderzoeksinformatie in hun eigen instellingssysteem (CRIS) open te stellen. Dit is een essentiële randvoorwaarde voor realisatie van ORIA.

NB. Het CRIS van instellingen is vaak verrijkt met gegevens van commerciële partijen⁸ (bv Scopus van Elsevier). Deze informatie mogen de instellingen ook publiekelijk delen⁹. Met betrekking tot andere commerciële leveranciers van verrijkte informatie (in een CRIS) moet het gesprek nog worden gevoerd.

2.2 Kwaliteitsimpuls

Organiseren van een (zoveel mogelijk automatische) feedbackloop naar gebruikte bronnen van onderzoeksinformatie (o.a. CRISsen, databases en repositories) voor volledige, betrouwbare en herbruikbare informatie. Hiertoe dienen ook processen en workflows uitgewerkt en geoptimaliseerd te worden. Deze terugkoppeling vanuit centrale (inter)nationale bronnen naar verbeterde kwaliteit (lokale) bronnen zal iteratief ontwikkeld moeten worden binnen diverse (pilot)projecten gekoppeld aan één of meerdere usecases.

2.3 Pilots en projecten

Pilots en projecten zijn nodig om functionaliteiten te demonstreren, voor zoals in paragraaf 1.1 besproken in punten A-G, op basis van (her)bruikbare onderzoeksinformatie.

Sommige pilots en projecten werden en worden met partijen als Elsevier, Dimensions en CWTS uitgevoerd. Het principe van een centrale OKB is in 2020 uit het contract met Elsevier gehaald (zie bijlage 6) maar binnen het contract worden er wel pilots gedaan die ook onderzoeksinformatie

⁶ Systemen waarin we resultaten van wetenschappelijk onderzoek vastleggen, noemen we CRIS, Current Research Information Systems. Deze systemen komen we tegen bij instellingen die aan wetenschappelijk onderzoek doen.

⁷ KUOZ = Kengetallen Universitair Onderzoek

⁸ Je zou ook kunnen stellen dat de Scopus of WoS data voor instellingen het uitgangspunt vormen voor de informatie en dat lokale CRISsen hier verrijkingen op doen. Aangezien we uitgaan van het beschikbaar stellen van beide informatiestromen, maakt dat voor de argumentatie hier geen verschil.

⁹ Elsevier heeft aangegeven dat instellingen informatie uit Scopus open mogen delen zolang het hun eigen output betreft. De vraag is of we dit nog moeten formaliseren of dat de schriftelijk bevestiging samen met de voorwaarden van Scopus voldoende is.

verbeteren en verrijken (zie bijlage 10). Elsevier ontwikkelt op deze wijze standaard diensten, de instellingen krijgen op deze wijze verbeterde en verrijkte onderzoeksinformatie terug naar hun systemen. Het is essentieel dat ook de informatie uit deze pilots door de instellingen open wordt gesteld.

Het model met Elsevier lijkt op het API-standaarden model zoals beschreven in de haalbaarheidsstudie door Dialogic met de daarin beschreven mogelijke risico's aangaande herbruikbaarheid en openheid. Voor een nationale ORIA is het van belang dat publieke partijen een gedegen afweging maken welke pilots en projecten men via een private partij wilt laat lopen en welke daarbuiten gehouden moeten worden.

De pilots en projecten binnen ORIA moeten de zorgen zoals samengevat in 1.2 t/m 1.5 moeten adresseren. Om de meerwaarde van ORIA in 2023 verder aan te tonen zullen een aantal gerichte pilots en projecten gestart moeten worden. Deze zullen informatie opleveren over:

1. hoe publieke instellingen hun onderzoeksinformatie (in CRIS) zelfstandig kunnen verrijken gebruik makend van open bronnen
2. hoe we de informatie in lokale onderzoeksinformatiesystemen verder kunnen uitbreiden met nieuwe informatie op basis van gezamenlijke afspraken en andere bronnen.
3. hoe met open databronnen zo volledig en betrouwbaar mogelijk informatie gegenereerd kan worden;
4. hoe een centrale database/datagraph gebaseerd op open bronnen én de lokale informatiesystemen kan verbeteren én op nationaal/instellingsniveau overzichten kan generen met verbeterde kwaliteit (bv. ontgaan van dubbelingen).

Een beperkt aantal pilots en projecten zullen in 2023 niet alleen de functionaliteiten demonstreren maar ook mogelijkheden en onmogelijkheden op bovenstaande vragen. In de loop van 2023 mede zal o.b.v. uitkomsten pilots en projecten een vervolgvorstel bestuurlijk worden voorgelegd om ORIA al dan niet verder op te schalen inclusief organisatiestructuur, rekening houdend met uitfasering NARCIS en zal een additionele impulsfinanciering bij NWO Regieorgaan Open Science ingediend worden.

Voor 2023 wordt gedacht aan

- aan project rond het kunnen monitoren van andere output dan alleen peer reviewed artikelen (bv. boekhoofdstukken, boeken en conference proceeding gaat) in context van vernieuwde open access monitor
- de meerwaarde van een PID Graph op het gebied van (open) CRIS data

Beslissing over de PID Graph pilot ligt eind november voor bij Raad van Bestuur NWO (zie bijlage 9). De beslissing over andere pilots zal begin 2023 genomen moeten worden door stuurgroep ORIA mee o.b.v. beschikbare budgetten (deel hangt af van bijdragen partijen buiten UNL).

Vervolg pilots en projecten vanaf 2024 zouden zich kunnen richten op andere use-cases (A-G), bijvoorbeeld gericht op:

- nieuwere en rijkere indicatoren en kwalitatieve narratives ter ondersteuning vernieuwing van research assessment en open science en;

- informatie over andersoortige output van hoger onderwijs (bv metadata van en PIDs voor onderzoeksdata, praktijkgericht onderzoek, software, leermaterialen, bijdragen in media, beleid, patenten en startups enz.).

dan wel gerichte pilots op basis van ervaringen opgedaan in 2023.

Andere voorbeelden van andere mogelijke pilots en projecten die eerder zijn genoemd zijn o.a.

- jaarlijkse monitor PhD proefschriften,
- invoeren van instellings-ID,
- relatie met andere repositories en eventueel register nevenwerkzaamheden hoogleraren.
- relatie leggen met de Research Software Directory van eScience center als output van onderzoek
- borgen van huidige functionaliteit van NARCIS (resolver, harvester) bij de uitfasering van NARCIS bij DANS-KNAW in relatie tot een mogelijke centrale "ORIA-infrastructuur".

2.4 Uitgangspunten ORIA

De voorgestelde ORIA geeft enerzijds de mogelijkheid om de transitie te realiseren van de huidige situatie naar een situatie waarin meer en kwalitatief betere onderzoeksinformatie open beschikbaar is. Anderzijds geeft de ORIA de mogelijkheid op iteratieve wijze toe te werken naar het nader te definiëren business model van de gezamenlijke investeringen in bijhorende "ORIA-infrastructuur" (diensten cf API-model dan wel centrale knowledge base/graph) met als uitgangspunten:

1. Deze ORIA infrastructuur gaat op termijn o.a. (enkele) NARCIS functionaliteiten vervangen in relatie tot aanleveren gegevens aan internationale databases. DANS heeft aangegeven NARCIS te willen uitfaseren en is al gestopt met doorontwikkeling ervan. Desalniettemin heeft DANS aangegeven een bijdrage te leveren aan doorontwikkeling van NARCIS functionaliteiten binnen ORIA.
2. Onderzoeksinformatie en analyses dienen open beschikbaar te zijn voor iedereen én van voldoende kwaliteit en betrouwbaar te zijn. In principe is dit soort informatie open tenzij er gewichtige bezwaren zijn vanuit AVG, veiligheid/bescherming academici inzake publieke opinie en overheidsinmenging, ethische aspecten, enz.
3. Het streven is om zoveel mogelijk en op termijn alleen maar open databronnen te gebruiken.
4. Een belangrijke bron van onderzoeksinformatie zijn de CRISsen van de instellingen. Deze dient dan ook zo snel mogelijk, en waar mogelijk, open beschikbaar te komen (wat deels al zo is i.r.t. NARCIS).
5. Correcties t.b.v. kwaliteit en volledigheid vanuit "ORIA-infrastructuur" altijd (zoveel mogelijk automatisch) terugvloeien naar de gebruikte databronnen zodanig dat fouten bij de bron gecorrigeerd worden
6. Deze informatie onder een open licentie publiekelijk beschikbaar, bij voorkeur met CC0 licentie.
7. Kennisinstellingen in hun contracten met externe partijen/met derden de Guiding Principles V2.1 implementeren zodanig dat (verrijkte) informatie over onderzoek terugvloeit naar het publieke domein, bij voorkeur met een CC0 licentie.
 - Uitgevers contractueel ertoe worden aangezet om de metadata over resultaten en projecten van Nederlandse onderzoekers te beschikbaar te stellen onder CC0 licentie, en daarmee ook in OKB wordt opgenomen.
 - Mogelijk geflankeerd door Europese wet- en regelgeving

3. Governance

Om de voorgestelde ORIA te realiseren wordt de volgende governance structuur voorgesteld met de kanttekening dat de structuur naar verloop van tijd verder aangescherpt c.q. uitgewerkt zal worden:

3.1 Bestuurlijke stuurgroep

De stuurgroep kent afvaardiging van een brede dwarsdoorsnede van nationale belanghebbenden, die beslist over gezamenlijke investeringen, de inhoudelijke koers bewaakt van de agenda, en de monitoring verzorgt van voortgang op hoofdlijnen. In beginsel zullen alle partijen die ORIA financieren deel uitmaken van de bestuurlijke stuurgroep.

3.2 Stuurgroep projectenbureau en projectenbureau

De uitvoering - van de door bestuurlijke stuurgroep vastgestelde ORIA - ligt bij een projectenbureau (zie bijlage 7) met stuurgroep. Om de vaart erin te houden zal het projectenbureau in 2023 bij SURF worden ondergebracht. In de eerste fase zal een voorstel worden ontwikkeld over n.t.b. operationele organisatie verantwoordelijk voor de uitvoering van ORIA ambities en jaarprogramma. Dit voorstel zal worden afgestemd met stuurgroep projectenbureau waaronder DANS (m.b.t transitie NARCIS), de betrokken kennispartijen (inclusief UKB en COS), en CSC-WO¹⁰ (m.b.t. de activiteiten rondom open onderzoeksinformatie).

Ook is er sprake van een projectorganisatie per project. Het projectenbureau bij SURF zal diverse projecten faciliteren en coördineren op samenhang en bijdragen aan ORIA.

3.3 Strategisch klankbordgroep

Een strategisch klankbordgroep wordt voorgesteld om (internationale) ontwikkelingen af te stemmen en vertaalslag te maken naar ORIA. Klankbordgroep is adviserend m.b.t. voorstellen pilotprojecten en jaarplannen ORIA.

4. Financiën

4.1 Budget en dekking

Voorstel is om eerst in 2023 een aantal pilots en projecten doen die de meerwaarde van ORIA aantonen (zie bij 3.3) en de investeringen beperkt te houden. Op basis van (inter)nationale ervaringen lijkt een jaarlijks budget van minimaal €700.000 noodzakelijk. Dit zal o.b.v. pilots en projecten in 2023 nader onderbouwd worden en bestuurlijk na de zomer 2023 voorgelegd worden. Voor 2023 is een totale budget van €530.000 nodig waarvan OCW via NWO €180.000 inbrengt. Het is de verwachting dat eind 2022 NWO-bestuur al akkoord kan gaan met financiering van het eerste ORIA-pilot (PID graph, zie bijlage 9) als opmaat voor mogelijk vervolgfianciering andere pilotprojecten. De resterende €350.000 zullen in 2023 door de betrokken kennispartijen/stakeholders gefinancierd te worden. Voorgesteld om daarbij een verdeelmodel te hanteren na rato van Basisfinanciering Rijksbegroting 2022, onderzoeksdeel. Dit verdeelmodel zal worden toegepast op zowel de verdeling tussen partijen als tussen instellingen onderling. Vanuit UNL zullen de universiteiten daartoe k€192 vrijmaken conform model in bijlage 8.

¹⁰ CSC-WO = Coördinerend SURF Contactpersonen / [CSC-wo | SURF.nl](https://www.csc-wo.nl)

Daarnaast zal naar additionele middelen gezocht worden via Regieorgaan Open Science i.o. Dit zal vooral impulsfinanciering zijn voor een aantal jaren dan wel voor specifieke pilotprojecten. NWO regieorgaan zal zeer waarschijnlijk ook eigen bijdragen van de kennisinstellingen vragen die worden gedekt uit die centrale middelen. Naast financiering pilotprojecten dient ook projectbureau gefinancierd te worden om de pilotprojecten te faciliteren en coördineren. Dit zal een aparte (deels) overlappende projectorganisatie vergen waarbij (voorlopig, voor 2023) SURF als penvoerder optreedt.

4.2 Bestedingsvoorstel 2023

Voor 2023 gaan we uit van minimale begroting voor ORIA voor projectenbureau en een eerste pilot. Deze worden financieel gedekt via bijdragen universiteiten (€192.000) en OCW-middelen via NWO Regieorgaan Open Science (€180.000 uit de 4M€ 2022 middelen). UNL zal andere stakeholders/kennispartijen verzoeken een evenredige financiële bijdrage te leveren van €158.000 (zie bijlage 8). Daarnaast zal projectenbureau een vervolgaanvraag indienen in 2023 bij NWO Regieorgaan i.o. (uit de 20M€/jaar middelen vanuit OCW). Deze extra middelen voor 2023 vanuit stakeholder/kennispartijen en OCW via NWO Regieorgaan zullen besteed worden aan additionele projecten en pilots (zie 3.2) en opschaling projectenbureau. In onderstaande tabel gaan we ervan uit dat NWO Regieorgaan Open Science minimaal 1 pilot financiert en dat andere kennispartijen ook bijdragen. Hiermee komt totale begroting op €530.000 voor 2023. De werkzaamheden en bestedingen worden na de zomer van 2023 geëvalueerd in kader van begroting voor 2024 ev.

Tevens worden de voorgestelde projecten voor ‘open onderzoeksinformatie’ op het gebied van ‘Open Science versterken’ bij SURF worden gekoppeld aan de ontwikkeling van ORIA.

Uitvoering (pilot)projecten	€298.200,-	(zie 3.3 voor mogelijke projecten en pilots, vanuit NWO Regieorgaan zal een eerste pilot gefinancierd worden, zie bijlage 9)
Projectenbureau	€231.800,-	(zie 4.2 en bijlage 7)
Jaarlijks totaal	€530.000,-	
NWO regieorgaan open science 2022-2023	-€180.000,-	initieel financiering project “PID graph” (bijlage 9); 2023 mogelijk additionele impulsfinanciering
Bijdrage stakeholders 2023	€350.000,-	* €190.000 door universiteiten * Bijdragen andere partijen (€160.000) begin 2023 na bestuurlijk verzoek UNL (zie bijlage 8 voor verdeelmodel)

5. Bijlagen

Aanvullende informatie is te vinden in de volgende bijlagen:

- Bijlage 1: Gesproken stakeholders jan-jul 2022
- Bijlage 2: Genodigden voor workshop 13 juni 2022
- Bijlage 3: Use cases voor een OKB
- Bijlage 4: Aandachtspunten
- Bijlage 5: Risicoanalyse
- Bijlage 6: Processtappen tot nu toe en het vervolg
- Bijlage 7: Projectenbureau bemensing
- Bijlage 8: Voorgesteld verdeelmodel financiële dekking 2023
- Bijlage 9: Projectvoorstel PID graph (zoals ingediend bij NWO voor financiering 1e pilot)
- Bijlage 10: Pilots binnen het contract met Elsevier

Bijlage 1: Gesproken stakeholders jan-jul 2022

- NWO
- UKB/UKBsis
- UKB Coordination Point Research Impact
- DANS-KNAW
- SURF CSC-WO
- TU Delft bibliotheek
- KB
- Elsevier stuurgroep publieke stakeholders
- NFU
- IDfuse
- Elsevier
- Clarivate
- JISC
- Research.fi / CSC Finland
- NORA research portal Denmark

Bijlage 2: Genodigden voor workshop 13 juni 2022

Naam	Instelling
Rik Iping	Erasmus MC
Nick Veenstra	TU Eindhoven
Alastair Dunning	TU Delft
Armand Guicherit	TU Delft
Marcel Broekarts	TU Delft
Maarten Steenhuis	KB
Marc van den Berg	KB
Astrid van Wesenbeeck	KB
Martijn Kleppe	KB
Inge Hofsink	KB
Iwan Holleman	CSC-WO
Jan-Willem Brock	CSC-WO
Hilde van Wijngaarden	UKB
Arjan Schalken	UKB
Lieuwe Kool	Amsterdam UMC
Guus van den Brekel	UMCG
Wim Hugo	DANS
Chris Baars	DANS
Wilko Steinhoff	DANS
Lambert Speelman	NWO
Hans de Jonge	NWO
Maarten Hoogerwerf	UU
Jeroen Sondervan	UU
Maurice Vanderfeesten	VU
Jeroen Bosman	UU
Bianca Kramer	UU
Clifford Tatum	CWTS
Magchiel Bijsterbosch	SURF
Sarah de Rijcke	CWTS
Ronald van Schijndel	Amsterdam UMC
Nanne De Vries	Maastricht University
Tom van Veen	SURF
John Doove	SURF
Wendy van Ginkel	Language in Interaction consortium
Tung Tung Chan	Erasmus Universiteit
Zohreh Zahedi	KNAW

Bijlage 3: Use cases voor ORIA

Uit gesprekken met stakeholders is geen éénduidig beeld van wat een Open Knowledge Base (OKB) zou moeten zijn. Wel is duidelijk dat er gezamenlijk gewerkt moet worden aan een Open Research Information Agenda (ORIA) waarbij de lokale CRISen en centrale OKB gevoed worden door een aantal pilotprojecten. De gesprekken met stakeholders hebben wel duidelijk gemaakt dat er voor verschillende gebruikersgroepen verschillende use cases (en functionaliteiten) zijn waar een OKB (al dan niet op termijn) in kan voorzien. Op basis van de gesprekken worden de volgende clusters van use cases onderscheiden. Deze kunnen als leidraad genomen worden voor de diverse pilotprojecten die de komende jaren uitgevoerd kunnen worden. Sommige (vele) pilotprojecten borduren voort op bestaande processen en uitvragen (zoals SEP, OA-monitor, KUOZ uitvraag, enz.). Andere pilotprojecten zullen gericht zijn op verbeteren van kwaliteit dan wel generen additionele overzichten. De onderstaande use cases voor een OKB kwamen nadrukkelijk naar voren in gesprekken met stakeholders en beslaan veel van de eerder in dit document genoemde functies van onderzoeksinformatie (zoals in paragraaf 1.1 besproken in punten A-G).

Use case 1: Rapporteren

Het kunnen rapporteren als individu, als instelling, en als koepels/land over onderzoek. Rapportage gegevens worden nu verzameld via verschillende invoeromgevingen zoals CRIS binnen instelling, NARCIS, de omgeving van uitgevers en van subsidiegevers. Voor goede betrouwbare rapportages wordt steeds meer vertrouwd op commerciële systemen als Web of Science (i.c.m. Unpaywall) en Scopus dan wel worden diensten afgenomen om eigen (CRIS) systeem te verbeteren.

Rapportages worden o.a. uitgevoerd ter verantwoording naar subsidiegever, en voor sturing (op impact). Bijv. wat je resultaten van onderzoek zijn geweest; aantal artikelen, % Open Access, tot aan (wetenschappelijk en maatschappelijk) impact van je onderzoek. Voor deze functie zijn volledigheid en betrouwbaarheid van data belangrijk.

Use Case 2: Evalueren en beoordelen

Evalueren is vooral formatieve assessment, en beoordelen vooral summatieve assessment. Evalueren en beoordelen geschiedt op individueel- (i.h.k.v. loopbaanontwikkeling), onderzoeksgroeps- (bv. SEP), en instellingsniveau.

Voor deze functie is het belangrijk dat er transparantie is over de gebruikte informatie voor evaluatie en beoordeling. Daarbij is er landelijk (en Europees) een beweging om tot rijkere data te komen en het belang van kwantitatieve data en (internationale) rankings te verminderen. Kwalitatieve gegevens (bv narratief) krijgen een groter belang naast de bredere (maatschappelijke) impact (bv via altmetrics). Zie ook 1.4

Use Case 3: Analyseren

Analyses worden uitgevoerd om tot (nieuwe) inzichten te komen over hoe onderzoek in Nederland wordt uitgevoerd. De analyse functie heeft ten doel inzicht te krijgen. Met behulp van tools en technieken kunnen overzichten gegenereerd worden. Deze overzichten en bijhorende conclusies daaruit kunnen gebruikt worden in communicatie met derden. Analyses bieden inzichten om op te sturen op bijvoorbeeld prioriteiten binnen het onderzoekslandschap en de resultaten van het onderzoek in het scholarly communication landschap. Om deze inzichten te vergroten en andere analysemogelijkheden te kunnen faciliteren is een 'level playing field' voor het gebruik van de onderzoeksinformatie van belang.

Voor de functie van analyse zijn daarnaast reproduceerbaarheid, transparantie en betrouwbaarheid van de analyse belangrijk.

Use case 4: Archiveren, opslaan en ontsluiten

Archiveren, opslaan en ontsluiten is het zodanig beheren en bewaren van informatie, dat deze nu en ook later bruikbaar is voor andere doelen. Hier hoort ook toegankelijk maken en vindbaar maken van al deze informatie. In deze context wordt sterker ingezet op het vindbaar maken en van onderzoeksproducten via metadata van die producten, dan op het aanbieden van relevante contextinformatie (zoals wie financiert het onderzoek, wat zijn de relaties tot ander onderzoek, en wat is de impact van het gevonden product). Deze use case borgt dan ook de duurzame toegang van de metadata en de (vindbaarheid van de gearhiveerde) objecten. Voor deze use case vervult de Koninklijke Bibliotheek een grote nationale rol en zijn traditioneel de instellingsbibliotheken de expertisehouders.

Recentelijk wordt ook steeds meer aandacht gevraagd voor de informatie over alternatieve output (zoals data, software, leermaterialen, etc.), het toepassen van standaarden op deze output (zoals FAIR), open infrastructuur, duurzame toegang en publieke sturing.

Use case 5: Informeren en dissemineren

Het brede publiek (bedrijven, citizen scientists, politici, journalisten, etc.) informeren over onderzoek in Nederland is een functie om 'societal engagement' en meer maatschappelijke impact te realiseren. Door onderzoeksinformatie zoals wie produceert welke onderzoeksoutput, in samenwerking met wie, en hoe is dit gefinancierd, beschikbaar te stellen kunnen burgers en bedrijven kennisnemen van de informatie en mogelijk een nieuwe toepassing vinden. Hierbij zal in toenemende mate verschuiving plaatsvinden van output aan het einde van de onderzoeksketen naar betrokkenheid eerder in de onderzoeksketen. Verder zullen er ook "producten" komen die maatschappelijke actoren makkelijker maken om wetenschappelijke kennis bruikbaar/begrijpelijk in te zetten (van overlays tot open education/open leermaterialen). Dat betekent dat er onderzoeksinformatie uit meer verschillende onderzoeksfases en over meer verschillende onderzoeksoutput beschikbaar zal komen.

Bijlage 4: Genoemde aandachtspunten in de business case studie

1. 'Open Knowledge Base' is een verwarrende titel. Er is een wens om voor een andere benaming te gaan passend bij de visie en functionaliteiten. Op basis van gesprekken met stakeholders en input van vele experts is gesuggereerd om de gezamenlijke Open Research Information Agenda (ORIA) te hanteren waarbij een centraal systeem een mogelijkheid zou kunnen worden.
2. Een gefaseerde aanpak is een veelgenoemde aanbeveling voor het ontwikkelen van een ORIA en daarbinnen voor een eventuele OKB door de stakeholders. Niet alle use-case kunnen direct opgepakt worden. Op een iteratieve wijze moet invulling gegeven worden aan de gezamenlijk agenda via gerichte pilots en projecten.
3. Hoger onderwijs heeft een referentie architectuur (HORA) en een doelarchitectuur (HOSA) en het is aan te bevelen dat de ORIA langs de lijnen van de HORA en de HOSA wordt uitgewerkt. Daarnaast zal de architectuur van ORIA (inclusief wat decentraal, wat via APIs, wat via een central database/datagraph) zal ontwikkeld en afgestemd moeten worden met internationale initiatieven als OpenAIRE, OpenAlex, Open citations, EOSC¹¹, enz.
4. Het is essentieel om nu al te starten met afspraken maken over uitwisseling, standaarden en identifiers ongeacht de vorm van een OKB. Dit is een sine qua non voor het borgen van publieke waarden zoals uitwisselbaarheid, beschikbaarheid en openheid van de informatie over onderzoek. Naaste generieke (architectuur) afspraken, zullen er ook afspraken gemaakt in de pilots en projecten zelf. Dus ook op afspraken niveau is er sprake van een iteratieve benadering.
5. Een centraal vraagstuk in Open onderzoeksinformatie is; hoe creëren we een meer level playing field voor (commerciële) partijen en hoe kunnen kennisinstellingen de publieke waarden en academische soevereiniteit waarborgen. Aanknopingspunten zijn te vinden in de implementatie [Guiding Principles 2.1](#) van landelijke Taskforce, of implementatie van het [LERU-data statement](#), de lobby binnen Europa (EUA) voor mogelijk Europese wet- en regelgeving (opgenomen als actie in ERA20, zie ook [EUA video Karen Maex](#) en [expert voice](#))
6. De informatie over onderzoek en bijbehorende analyses moeten van voldoende kwaliteit zijn omdat deze informatie steeds meer gebruikt wordt in rapportages (instellingen, koepels, landelijk) en gebruikt kan worden in verantwoorden impact (SEP tot aan beoordelen subsidieaanvragen). Bovendien is er in het kader van Erkennen & Waarderen noodzaak voor rijkere (kwalitatieve en kwantitatieve) research information. Extra aandacht voor kwaliteit en volledigheid van onderzoeksinformatie zijn dan ook van belang naast het openbaar beschikbaar maken van onderzoeksinformatie.

¹¹ EOSC = European Open Science Cloud

Bijlage 5: Risicoanalyse

<i>Id</i>	<i>Risico</i>	<i>Kans (1-5)</i>	<i>Omvang (1-5)</i>	<i>Maatregel</i>
1	Nationale oplossing compliceert of verdubbelt internationale oplossingen	4	3	Door als uitgangspunt te nemen dat institutionele en nationale bronnen leidend zijn voor ORIA en die uiteindelijk internationale bronnen moeten voeden, wordt op langere termijn dit risico afgedekt. Ook om in diverse pilots en projecten juist iteratief te werken met open databronnen die juist in internationale context (door)ontwikkeld worden.
2	Benodigde expertise is zowel binnen ORIA projecten en pilots als bij de instellingen is schaars, waardoor een te kleine groep te veel (nieuwe) inzet moet leveren	4	5	Door calls uit te zetten via pilots en projecten wordt een deel van de ontwikkelingen juist neergelegd bij degenen die enthousiast zijn en expertise/netwerken hebben. Door te linken aan CRISsen en informatiesystemen koepels wordt er geleund hun expertise/inzet. Verder wordt voortgeborduurd op de datahub van UKBsis (met dataexperts en netwerk externen). Aangezien ook de institutionele repositories onderzoeksinformatie bevatten die niet altijd mee wordt genomen in een CRIS worden ook expertise met betrekking van deze systemen betrokken
3	Pilots en projecten vergen ook inzet vanuit instellingen en o.a. CRIS-managers	5	3	Benodigde inzet vanuit de instellingen zal begroot moeten worden in elke pilot en project. Soms (deels) gefinancierd, soms als in-kind bijdragen. Deze inzet moet voor start van elke project / pilot gespecificeerd en akkoord bevonden zijn.
4	ORIA veroorzaakt concurrentie met marktpartijen i.p.v. samenwerking of complementariteit	2	2	Er zal kritisch gestuurd moeten worden op a) creëren level playing field marktpartijen en b) meerwaarde additionele services bovenop ORIA-infrastructuur. De services van verschillende partijen (publiek of privaat) kunnen een rol spelen bij verbeteren de informatie die nu lokaal in het CRIS staat en het van genereren verrijkte informatie dan wel andere (kwalitatieve) indicatoren.
5	ORIA versterkt/behoudt de (ongewenste) focus op kwantitatieve gegevens i.p.v. kwalitatieve gegevens	3	3	Naast basis use cases voor ORIA zullen er juist gegevens ontwikkeld moeten worden die bijdragen aan brede Open Science doelen i.r.t. maatschappij en die bijdragen aan het kwalitatieve verhaal rond onderzoek(simpact). Zie ook 1.4
6	De besluitvorming rond ORIA verloopt over te veel schijven en is daarmee te traag	4	4	Door de co-financieringsconstructie (voorzien via het NWO Regieorgaan Open Science i.o.) wordt een duidelijk tijdsplan opgesteld en is er druk om doelmatig en binnen de tijd en financiële kaders de projectresultaten op te leveren. Dit noopt tot efficiënte besluitvorming door de stuurgroep van o.a. mee betalende kennispartijen. Deze dient te zorgen voor snelle besluiten en acties. Door de financiering van

Id	Risico	Kans (1-5)	Omvang (1-5)	Maatregel
				meerdere stakeholders dienen er afspraken gemaakt te worden over hoe besluitvorming plaatsvindt.
7	De ontwikkeltijd van een centrale ORIA-infrastructuur wordt door scope creep te lang	3	4	Frame: ORIA vraagt continue doorontwikkeling. In het begin van project zal geborgd moeten worden dat op de lange termijn zowel voor een ORIA-infrastructuur zelf als voor de continue aanlevering van kwalitatief hoogwaardige informatie over onderzoek structureel een investering is. Op korte termijn zal gestuurd moeten worden op de successen uit pilots en projecten (investeringen in open onderzoeksinformatie tot hoger level tillen) en via agile trajecten aan uitbreiding van use-cases gewerkt worden. Prioritering vindt op stuurgroep niveau plaats in combinatie met bijbehorende financiën.
8	Onbedoelde uitsluiting van kleine commerciële partijen die niet/onmogelijk aan bepaalde voorwaarden kunnen voldoen	3	4	Een aantal van de te operationaliseren guiding principles is mogelijk moeilijker uitvoerbaar door startups en leveranciers van kleine omvang. Zoals de principes op governance door publieke stakeholders en open algoritmes (waar startende partijen hun plaats in de markt mee willen verwerven). Anderzijds leveren de Guiding Principles uiteraard ook mogelijke voordelen op voor die organisaties die juist al op basis van open standaarden werken. Per opdracht aan een externe leverancier moeten afspraken worden gemaakt in contracten ten aanzien van wat er terug naar het publieke domein (bijvoorbeeld het recept achter algoritmen) moet vloeien en wat er als meerwaarde van de marktpartij kan worden aangeboden als service (nieuwe indicator, antwoorden op impact vraagstukken etc.)
9	De markt heeft blokkademacht op het gebied van aanleveren van (betrouwbare) onderzoeksinformatie	4	5	Vele CRISsen van de universiteiten maken gebruik van Elsevier software (PURE). De gegevens in PURE van universiteiten zijn nauw verwant aan de data uit SCOPUS van Elsevier welke gegevens bevat van andere uitgevers. Het is niet geheel duidelijk waar eigenaarschap van deze gegevens liggen (Elsevier geeft steeds aan dat dit bij publieke partijen ligt, bij andere commerciële partijen kan dat anders liggen) en hoe dit zich verhoudt dat het databankrecht. Er moet gestuurd worden op de vaststelling dat eigenaarschap op metadata en afgeleide gegevens altijd in publieke domein moeten liggen. Ook wordt via Guiding Principles gestuurd op dat verrijkte data uit commerciële systemen open moeten zijn. Dit is wel een afspraak voor pilots binnen het contract met Elsevier

Id	Risico	Kans (1-5)	Omvang (1-5)	Maatregel
				<p>(maar die vloeit terug naar CRIS waarbij instelling deze nog moet openstellen) maar dat is nog niet zo voor andere commerciële service providers. Wel heeft Elsevier aangegeven dat data in Scopus van andere uitgevers door de instellingen open gedeeld mag worden zolang het hun eigen output betreft.</p> <p>Dit risico wordt op lange termijn tegengegaan door de keuze om (op termijn) alleen met open databronnen te gaan werken en door steeds vergaande afspraken met commerciële partijen te maken, liefst in Europees verband. De financiering van pilots en projecten is ook bedoeld om het het werken met open databronnen te versnellen en om toe te groeien dat we uiteindelijk alleen maar met open databronnen te gaan werken.</p>
10	Impact van Europese wetgeving bemoeilijkt delen van open onderzoeksinformatie (bv. Digital Services Act)	3	4	<p>Er is al een sterke lobby gaande voor goede Europese wet- en regelgeving die de positie van universiteiten juist moet versterken (al duurt dat >5 jaren).</p> <p>Aan de andere kant zorgt de Digital Services Act voor verzwaarde maatregelen die elke platform moet treffen bij 'misinformatie' en monitoring daarvan. Dit wordt in begroting van centrale coördinatie meegenomen.</p>
11	Uitfaseren NARCIS loopt sneller dan realiseren OKB waardoor functionaliteit verloren dreigt te gaan	4	4	<p>DANS actief betrekken bij ontwikkeling van "ORIA-infrastructuur" en mede zorgdragen dat voldoende tools/use cases gereed zijn voordat NARCIS eruit gaat. DANS heeft aangegeven welwillend te zijn in meedenken en meewerken in transitietrajecten.</p> <p>Stuurgroep zal in de uitwerking actief moeten sturen.</p> <p>https://www.narcis.nl/dataprovider/Language/nl</p>
12	"Garbage in, garbage out" in het bronsystemen zoals het CRIS bij de instellingen	4	3	<p>Feedbackloop naar de bronsystemen en investeringen in verbeteren bronsysteem inclusief CRISsen (zie 2.5). Provenance moet duidelijk zijn als hoofdvoorwaarde, en verder blijft het dan aan de instellingen of ze het OKB als bron voor verrijkte data accepteren of niet, en hoe ze omgaan met evt. conflicterende informatie. PROV Model Primer (w3.org)</p> <p>O.a. afspraken en standaarden vastleggen rondom datakwaliteit en actualiteit</p>
13	Parallele en daardoor mogelijk divergerende ontwikkelingen rondom andere	4	2	<p>Organisaties die richten op andere onderzoeksoutput dan artikelen (software, leermaterialen, FAIR datasets) zullen aangehaakt moeten worden zodanig dat hun directories (metadata-databases) gekoppeld kunnen worden aan</p>

<i>Id</i>	<i>Risico</i>	<i>Kans (1-5)</i>	<i>Omvang (1-5)</i>	<i>Maatregel</i>
	research objecten en PIDs			datahub OKB. Dit kan door cofinanciering in de pilots en projecten maar ook door hen te betrekken in ontwikkelplan datahub op centraal niveau.
14	De focus op CRIS data leidt tot het verlies van waardevolle onderzoeksinformatie uit andere bronnen	3	3	In een vroege fase van de beoogde pilotprojecten wordt een architectuurschets opgesteld van de verschillende bronnen waar onderzoeksinformatie over Nederlands onderzoek zich bevindt. Zowel internationale bronnen (zoals; Unpaywall, BASE, OpenAlex, OpenAIRE) als de nationale en lokale bronnen (zoals; NARCIS, de institutional repositories, data repositories, de Research software directory). Dat moet ervoor zorgen dat (al dan niet gefaseerd) de relevante bronnen van informatie in scope zijn en blijven.
15	Door het combineren van data die los van elkaar geen privacy issue oplevert, kan er plots wel privacy issues ontstaan. bijv. (zeer) transparant maken welke onderzoeker niet funder compliant werkt.			Op dit moment kunnen we moeilijk inschatten welke informatie privacy en/of veiligheid van academici schaadt. In het projectenbureau ligt er een rol voor de jurist om niet alleen te toetsen welke informatie e.d. de projecten benutten, maar ook de uitkomsten van de projecten in het kader van analyses en nieuwe inzichten.

Bijlage 6: Processtappen (tot nu toe en het vervolg)

Proces tot nu toe

- Eind 2019 OKB vermeld in Framework contract met Elsevier
- Begin 2020 is een landelijke Taskforce ingesteld die condities vanuit publieke partijen heeft geformuleerd voor research information/analysis
- Op verzoek van die Taskforce is OKB in mei 2020 uit contract met Elsevier gehouden hangende besluitvorming publieke partijen
- Dialogic heeft in opdracht van die landelijke Taskforce, gefinancierd vanuit OCW, een haalbaarheidsstudie uitgevoerd
- In april 2021 is die haalbaarheidsstudie in NPOS stuurgroep besproken met vervolgens verzoek bestuurlijk draagvlak te peilen
- UNL, NWO, ZonMW, KNAW, SURF en VH geven aan dat er bestuurlijk draagvlak is voor ontwikkeling van een OKB zoals besproken in NPOS stuurgroep juni 2021
- In september 2021 heeft Magchiel Bijsterbosch de principes en ideeën achter OKB in NPOS stuurgroep gepresenteerd. Stuurgroep geeft aan dat een businesscase studie nodig is
- In december 2021 heeft UNL-bestuur deze handschoen opgepakt en aan SURF verzocht om samen de business case studie uit te voeren t.b.v. alle bij NPOS betrokken kennispartijen en koepels (UNL, NWO, NFU, SURF, KNAW, VH, DANS en ZOnMW)
- Gedurende de periode jan-jul is met vele stakeholders gesproken (zie bijlage 1) en een workshop georganiseerd met 35 experts (zie bijlage 2).
- Verkrijgen detail terugkoppeling van de 35 betrokken experts op diverse versies tussen augustus-oktober 2022
- Versie 0.9 geagendeerd bij NPOS Stuurgroep en Adviesraad eind september 2022 voor verkrijgen terugkoppeling op hoofdlijnen
- Verkrijgen detail terugkoppeling van betrokken partijen via landelijke Open Access afstemmingsoverleg in oktober-november 2022
- Bijgestelde versie naar Chiefs Open Science (2 november) en UKB (UNL-proces)
- Agendering bij bestuurlijke trekkers bij UNL (9 november) en bestuurlijke trekkers bij SURF

Vervolgproces

- Agendering bij rectoren UNL (SOO) d.d. 2 december 2021 inclusief bijbehorende investeringen
- Eventueel ook al bij NPOS stuurgroep en Adviesraad december, dan wel als separate voorstellen bij besturen van elk betrokken kennispartij.
- Separaat voorstel eerste pilot voor cofinanciering OCW middelen via NWO Regieorgaan in oprichting (bijlage 9)
- Voorkeur bestuurlijke route: na positief besluit UNL inclusief bijbehorende investeringen, verzoek naar andere kennispartijen om ook bij te dragen (als proces volgorde gaat i.p.v. parallel).

Bijlage 7 : Projectenbureau bemensing

Om de projecten binnen ORIA te faciliteren en te coördineren zijn een aantal rollen en expertises opgenomen. Dit zijn complementaire expertises die de projecten in goede banen kunnen leiden en de voortgang op ORIA kunnen bewaken.

Projectoffice (projectmanager en -support)	€73.200,-	600 uur	<ul style="list-style-type: none">- Bewaking/monitoring voortgang projecten- Rapportages aan stuurgroepen- Faciliteren projecten en werkgroepen
Architect/informatiespecialist	€73.200,-	600 uur	<ul style="list-style-type: none">- Samenhang van processen, informatie, etc. tussen projecten en met (inter)nationale initiatieven- Ondersteuning projectstartarchitectuur van projecten
Jurist	€36.600,-	300 uur	<ul style="list-style-type: none">- Juridische advisering m.b.t. gebruik van data, software, infrastructuren, etc.
Communitymanager	€36.600,-	300 uur	<ul style="list-style-type: none">- Communicatie over ontwikkelingen binnen ORIA
Communicatie	€12.200,-	100 uur	<ul style="list-style-type: none">- Communicatie uitingen ORIA- Ondersteuning bij website, social media, etc.

Bijlage 8: Voorgesteld verdeelmodel financiële dekking 2023

De Open Research Information Agenda (ORIA) en bijhorende (centrale) diensten/infrastructuur wordt gezien als een gezamenlijke investering voor een nationale infrastructuur. Het is essentieel dat dit gestuurd en gedragen wordt door alle publieke kennispartijen. Van de volgende partijen wordt daarom een financiële bijdrage gevraagd waarbij we een verdeelmodel hanteren naar grootte basisfinanciering rijksbegroting, deel onderzoek dan wel ontwerp en ontwikkeling. Onderstaande tabel is indicatief en moet aangepast worden naar begrotings-kengetallen die voor 2022 gelden.

Het betreft de financiële dekking voor 2023 met €350.000 van onderstaande partijen aangevuld met €180.000 door OCW via NWO Regieorgaan i.o.

ORIA-OKB financiering - bijdragen partijen	Basisfinanciering Rijksbegroting 2021	Bijdrage aan begroting 2023
UNL: Universiteiten (onderzoeksdeel WO)	€ 2.193.737	€ 192.065
NWO en ZonMW	€ 866.505	€ 75.864
NFU: Academische ziekenhuizen	€ 707.959	€ 61.983
KNAW	€ 87.882	€ 7.694
VH: Hogescholen (ontwerp en ontwikkeling hbo)	€ 91.757	€ 8.033
KB	€ 49.800	€ 4.360

ORIA start zodra UNL akkoord is waarna UNL-bestuur andere partijen zoals hierboven genoemd verzoekt bij te dragen. Daarbovenop wordt een impulsfinanciering vanuit NWO Regieorgaan Open Science i.o. verwacht (bijlage 9) en zal nog nader met DANS besproken moeten worden over hun bijdragen in kader van uitfasering NARCIS.

De bijdragen van de universiteiten (€192.065) zal conform hetzelfde verdeelmodel worden gehanteerd. Op basis van kengetallen begrotings-kengetallen 2022 is de onderverdeling universiteiten indicatief:

	Rijksbijdrage Onderzoek 2022 (1e brief)		Universiteiten
Universiteit Leiden	€ 198.371.956	8,77%	€ 16.844
Rijksuniversiteit Groningen	€ 204.043.174	9,02%	€ 17.324
Universiteit Utrecht	€ 254.624.074	11,26%	€ 21.627
Erasmus Universiteit Rotterdam	€ 134.490.640	5,95%	€ 11.428
Technische Universiteit Delft	€ 267.156.305	11,81%	€ 22.683
Technische Universiteit Eindhoven	€ 152.843.612	6,76%	€ 12.984
Universiteit Twente	€ 128.407.219	5,68%	€ 10.909
Wageningen University	€ 148.616.339	6,57%	€ 12.619
Universiteit Maastricht	€ 122.462.740	5,41%	€ 10.391
Universiteit van Amsterdam	€ 225.450.831	9,97%	€ 19.149
Vrije Universiteit Amsterdam	€ 178.588.384	7,89%	€ 15.154
Radboud Universiteit Nijmegen	€ 166.108.972	7,34%	€ 14.098
Tilburg University	€ 61.138.110	2,70%	€ 5.186
Open Universiteit	€ 19.799.902	0,88%	€ 1.690
	€ 2.262.102.258	100,0%	€ 192.065

Bijlage 9: Projectvoorstel PID graph (zoals ingediend bij NWO voor financiering 1e pilot)

Open Knowledge Base – PID Graph 2022 - 2023

Werkplan Een eerste pilot voor een OKB

Gül Akcaova, Darco Jansen, John Doove
20-10-2022

Managementsamenvatting

Naar aanleiding van onderzoek naar een business case voor een Open Knowledge Base wordt in dit werkplan een beschrijving gegeven voor een eerste pilot om uit te vinden welke use cases er kunnen worden bediend met een PID (Persistent Identifier) Graph. Daarnaast wordt onderzocht de PID Graph gebruikt kan worden om de kwaliteit van onderzoeksinformatie in lokale bronnen te verbeteren. Een PID Graph geeft inzicht in de beschrijvingen (metadata) en onderlinge relaties van onderzoeksobjecten (publicaties, onderzoekers, onderzoeksprojecten, onderzoekinstellingen) en hoe open en herbruikbaar deze zijn. De pilot leert ons welke toepassingen op welke termijn op basis van open onderzoeksinformatie mogelijk zijn en welke andere stappen vereisen. Beoogd wordt dat deze pilot wordt opgevolgd door andere pilots. Deze serie pilots is onderdeel van een bredere nationale open onderzoeksinformatie agenda.

Aanleiding en achtergrond

Om digitale soevereiniteit en kwaliteit te borgen op rapportage, evaluatie en vindbaarheid van Nederlands onderzoek is op verzoek van UNL een traject gestart waarin een business case is opgesteld voor een 'Open Knowledge Base' (OKB) voor informatie over onderzoek. Uit dit traject, gecoördineerd door SURF en UNL is gebleken dat de behoefte aan open onderzoeksinformatie breed wordt gedragen. Dit is bevestigd in een workshop met een groep van nationale belanghebbenden in juni 2022. Uit de studie is gebleken dat er consensus over de open benadering bestaat en dat via diverse projecten iteratief toegegroeid moet worden naar open beschikbare onderzoeksinformatie.

Ten eerste zijn er afspraken nodig die garanderen dat we in Nederland (en internationaal) dezelfde definities (denk bijvoorbeeld aan Open Access) en standaarden (denk aan het gebruik van PIDs) hanteren. Alleen dan is het aggregeren van onderzoeksinformatie over instellingsgrenzen heen mogelijk en kan deze open informatie zinvol worden gebruikt voor nationale inzichten, rapportage en evaluatiedoeleinden. Bovendien moeten we deze afspraken ook bij de marktpartijen kunnen afdwingen waar dat nog niet het geval is, als we gebruik willen maken van (verrijkte) data die in systemen staat opgeslagen die door marktpartijen worden geleverd. Ten tweede zijn niet alle brondata open en momenteel nog van onvoldoende kwaliteit en onvoldoende volledig om alle gewenste toepassingen, die in de OKB-business case zijn beschreven, van open onderzoeksinformatie te realiseren. Een kwaliteitsimpuls is hier nodig om te zorgen dat én de instellingssystemen (en andere open bronnen) én een Open Knowledge Base op termijn ook kwalitatief goede informatie kunnen bieden.

Om te verkennen welke toepassingen we met bestaande lokale bronnen en beschikbare open onderzoeksinformatie al kunnen faciliteren en welke niet, zijn meerdere (pilot)projecten nodig. Deze zullen iteratief bijdragen aan een groeiende OKB en tegelijkertijd de kwaliteit van instellingssystemen en andere open bronnen verbeteren. Ook zal duidelijk worden welke afspraken er moeten worden gemaakt of verscherpt over het aanleveren en gebruik van onderzoeksinformatie. Om deze reden wordt in de businesscase geconcludeerd dat er gezamenlijk naar een Open Research Information agenda toegewerkt moet worden, die zowel een serie pilotprojecten, een iteratieve aanpak naar een OKB en

een afsprakenstelsel omvat. Deze subsidieaanvraag is onderdeel van die brede onderzoeksinformatieagenda en richt zich op een eerste pilotproject; een Open Knowledge PID Graph.

PID Graph

Een PID Graph verbindt verschillende onderzoeksobjecten in het onderzoekslandschap op basis van PIDs en de bij die objecten behorende metadata. Deze onderzoeksobjecten kunnen bijvoorbeeld publicaties, leermaterialen, datasets of onderzoekssoftware zijn, maar ook docenten, onderzoekers, onderzoekinstellingen, -subsidies of -projecten. Deze pilot bekijkt wat er mogelijk is met die objecten waar nu al een PID voor aanwezig is. Op langere termijn moet er geïnvesteerd worden in afspraken dat elke object een PID heeft.

PID Graphs zijn in verschillende internationale projecten en organisaties zoals in OpenAIRE en FREYA opgesteld en zijn een goede manier om inzicht te krijgen welke objecten PIDs hebben, welke metadata bij deze objecten worden meegegeven en welke relaties er tussen de verschillende objecten bestaan. Een dergelijke PID Graph leert ons wat de mogelijkheden zijn voor toepassingen van deze open onderzoeksinformatie, zoals hoe het huidige databases kan verrijken en verbeteren. Ter illustratie; een PID Graph kan inzicht geven in alle output die kan worden herleid naar een bepaalde onderzoekssubsidie, kan instellingen helpen bij het bepalen van het percentage open access van onderzoeksoutput en onderzoekers kunnen gemakkelijke verbanden tussen onderzoeksprojecten vinden. Hieronder worden de meer specifieke doelstellingen van de pilot van voor de PID Graph uitgewerkt, alvorens de activiteiten die tot deze doelstellingen moeten leiden worden beschreven inclusief de daarvoor benodigde middelen.

Doelstellingen

1. Inzicht in ontwerpcriteria OKB en eisen voor onderliggende primaire bronnen
 - a. Inzicht in kwaliteit, volledigheid en (her)bruikbaarheid van bestaande lokale en/of nationale en internationale primaire en open databronnen
 - b. Inzicht hoe we de kwaliteit van de primaire bronnen zoveel mogelijk automatische kunnen verbeteren op basis van inzichten uit PID Graph
 - c. Ervaring met de techniek van een PID Graph
 - d. Inzicht in de verhouding tussen een nationale PID Graph en lokale en internationale bronnen (positionering in informatiearchitectuur)
2. Een community faciliteren die kennis opbouwt en deelt over open onderzoeksinformatie
3. Inzicht in benodigde financiën en sturing voor een nationale PID Graph
4. Waardepropositie van een PID Graph (toepasbaarheid voor use cases uit de business case)

Activiteiten

Om de doelstellingen te behalen zijn de volgende activiteiten gedefinieerd:

Doelstelling 1. Inzicht in ontwerpcriteria OKB

Resultaten	Hoe	Kosten	Matching
Inzicht in kwaliteit, volledigheid en (her)bruikbaarheid van bestaande lokale en/of nationale en internationale open databronnen	<p>Maken overzicht van informatieobjecten waarbij de informatie niet volledig is in de CRISsen en andere (inter)nationale bronnen.</p> <p>Identificeren wat de relaties zijn tussen de onderzoeksobjecten (met een PID), en waar de relaties nog niet zijn gelegd.</p> <p>Ontwerpen data/informatiemodel voor de PID Graph en voor een mogelijke OKB.</p>	<p>Projectmanager, € 6,000</p> <p>Architect, € 18,000</p> <p>Jurist, € 9,000</p> <p><u>€ 33.000,-</u></p>	Ten minste drie instellingen zullen in kind investeren om de PID Graph te realiseren en daarbij inzichten genereren en verzamelen voor de bruikbaarheid (beschikbaarheid, kwaliteit, volledigheid) van bronnen

Resultaten	Hoe	Kosten	Matching
Ervaring met de techniek van een PID Graph	<p>Voor ervaringen en geleerde lessen contact opnemen met internationale partners (RDA Open Science Graphs, TAPIR (tib), etc.).</p> <p>Inventariseren hoe makkelijk/moeilijk databronnen aangesloten kunnen worden.</p>	<p>Projectmanager, € 5,000</p> <p>Architect, € 15,000</p> <p>Ontwikkelkosten, € 20,000</p> <p><u>€ 40.000,-</u></p>	Ten minste drie instellingen zullen in kind investeren om de PID Graph te realiseren en daarbij ervaring op doen met de techniek van het maken van een PID Graph.
Inzicht in de verhouding tussen een nationale PID Graph en lokale, nationale en internationale bronnen (positionering in informatiearchitectuur)	<p>Afstemmen FAIRCORE4EOSC, OpenAIRE, etc.</p> <p>Schetsen van de verschillende bronnen, de onderlinge verhouding tussen de functionaliteit van deze bronnen en inzicht in de huidige koppelingen.</p>	<p>Projectmanager, € 5,000</p> <p>Architect, € 14,000</p> <p><u>€ 19.000,-</u></p>	SURF werkt aanvullende op deze samen met (Knowledge Exchange, ARDC, Research.fi, NORA) en participeert in (FAIRCORE4EOSC, RDA) verschillende internationale projecten waarin de verhouding tot internationale ontwikkeling wordt onderzocht

Doelstelling 2. Een community die kennis opbouwt en deelt over open onderzoeksinformatie

Resultaten	Activiteit	Kosten
Community building	Opstarten en inrichten communitymanagement. Contentcreatie en delen van ontwikkelingen, ideeën, en kennis met bredere community (bijv. via communities.surf.nl).	Communitymanager, € 10,000 Communicatie, € 7,000 <u>€ 17.000,-</u>
Community management	(Open) bijeenkomsten organiseren om met verschillende stakeholders te spreken over voortgang en vervolgstappen gedurende pilot.	Communitymanager, € 10,000 Kosten bijeenkomsten € 1.000 <u>€ 11.000,-</u>

Doelstelling 3. Inzicht in benodigde financiën en sturing voor een nationale PID Graph

Resultaten	Activiteit	Kosten
Governance	Vaststellen welke partijen minimaal zeggenschap hebben als “eigenaren” van brondata, een iteratieve Open Knowledge Base en van de pilots.	Projectmanager, € 5,500 Communitymanager, € 4,500 Jurist, € 9,000 <u>€ 19.000,-</u>
Financieringsmodel	Inschatting kosten voor continuïteit PID Graph incl. Doorontwikkeling.	Projectmanager, € 5,000 Communitymanager, € 4,000 <u>€ 9.000,-</u>

Doelstelling 4. Waardepropositie van een PID Graph (toepasbaarheid voor use cases uit de business case)

Resultaten	Activiteit	Kosten	
Definiëren gebruikersgroepen	Vaststellen voor welke use cases uit de business case de PID Graph bruikbaar is. Verkennen voor welke vragen PID Graph toepasbaar is en user stories vaststellen.	Projectmanager, € 5,000 Communitymanager, € 4,000 Communicatie, € 6,000 <u>€ 15.000,-</u>	Ten minste drie instellingen zullen in kind naar schatting 40 uur per instelling investeren om de PID Graph te realiseren en te toetsen voor welke use cases de PID Graph bruikbaar is of kan zijn.
Functionele wensen en eisen gedurende pilot	Eisen en wensen in kaart brengen en deze vertalen naar roadmap.	Projectmanager, € 5,000 Architect, € 8,000 Communitymanager, € 4,000 <u>€ 17.000,-</u>	

Resources (geconsolideerd)

Rollen	Totaal aantal uren	Kosten
Projectmanager	300 uur	36,5K
Communicatie	95 uur	13k
Jurist	150 uur	18k
Communitymanager	300 uur	36,5K
Architect	450 uur	55K
Ontwikkeling + overig out of pocket		21K
Totaal		180K

Overzicht aangevraagde financiering

1 november 2022 - 1 juli 2023

Inzicht in ontwerpcriteria OKB o.b.v. PID Graph	€ 92.000
Een community die kennis opbouwt en deelt over open onderzoeksinformatie	€ 28.000
Inzicht in benodigde financiën en sturing voor een nationale PID Graph	€ 28.000
Waardepropositie van een PID Graph (toepasbaarheid voor use cases uit de business case)	€ 32.000
Totaal	€ 180.000

Context voor de subsidieaanvraag

Zoals eerder gesteld wordt de beweging naar open research information breed ondersteund. Een werkgroep, bestaande uit medewerkers van verschillende onderwijsinstellingen, DANS-KAW en SURF die zich wil gaan richten op de PID Graph staat in de startblokken en draagt nu en de komende maanden in-kind bij aan de PID Graph. Daarnaast wordt momenteel onder coördinatie van SURF samen met universiteitsbibliotheken (Coördination Research Impact point), DANS-KNAW, NWO en CWTS gewerkt aan een nationale roadmap voor het gebruik van PIDs. Deze activiteiten versterken de PID Graph en vice versa. Tot slot wordt de hierboven beschreven bredere open research information agenda ook verder uitgewerkt en bestuurlijk voorgelegd (SOO van UNL bijvoorbeeld). De PID Graph jaagt daarmee de ontwikkelingen binnen open research information aan.

6. Verantwoording

De verantwoording verloopt uiteraard naar de subsidieverstrekker. Indien de financiering verloopt via een subsidiebrief van NWO aan SURF zal de verantwoording van SURF aan NWO qua planvorming via dit werkplan geschieden en qua rapportage via een daarvoor opgestelde rapportageformat in 2023. OKB-PID Graph zal uiteraard de bestedingen aanwenden binnen de kaders van de financieringsvoorwaarden. OKB-PID Graph zal zich daarnaast laten adviseren door experts en stakeholders uit het Open Science ecosysteem.

Bijlage 10: Pilots binnen het contract met Elsevier

Binnen het contract met Elsevier (2020-2024) wordt pilotsgewijs gewerkt aan de ontwikkeling van nieuwe diensten die vooral gericht zijn op verbeteren en verrijken van onderzoeksinformatie, en breder de Open Science ambities van Nederland ondersteunen. Zie gezamenlijke website:

<https://epdos.nl/nl/open-science-services/>

Momenteel lopende de volgende pilots:

- Website VU-UvA voor presentatie onderzoekers en de onderzoeksresultaten met samengevoegde en ontdubbelde data
- Datamonitor voor het vindbaar en toegankelijk maken van onderzoeksdatasets
- Grants monitor voor verbeteren van de open data flow van onderzoeksinvesteringen
- Telescope - RI Tracking om te testen hoe het gebruik van zogeheten Research Infrastructure (RI) in wetenschappelijke literatuur kan worden geïdentificeerd.
- Rare Diseases Analytics om publieke toegang te verschaffen tot onderzoek naar zeldzame ziektes in Nederland

De afspraak is dat de instellingen eigendom hebben en houden op de ingebrachte en verrijkte informatie. Dit conform voorwaarden zoals vastgelegd in contract als collaboration principes, gebaseerd op Guiding Principles V1.0 van de eerder genoemde Taskforce. Elsevier ontwikkelt op deze wijze standaard diensten, de instellingen krijgen op deze wijze verbeterde en verrijkte onderzoeksinformatie terug naar hun systemen (CRISsen). Het is dus aan de instellingen of men de gegevens uit deze pilots open beschikbaar stelt.

Voor een open research information agenda is het essentieel dat ook de informatie die in deze projecten wordt vastgelegd open wordt gesteld. Het model met Elsevier lijkt op het API-standaarden model zoals beschreven in de [haalbaarheidsstudie door Dialogic](#) met de daarin beschreven mogelijke risico's aangaande herbruikbaarheid en openheid (o.a. op het gebied van algoritmes en software).

Verder dienen we minder afhankelijk te worden van particuliere ondernemingen voor het verstrekken en verrijken van gegevens over onderzoek en bijhorende software. Daarbij dienen we vendor lock-in te voorkomen en een gelijk speelveld te creëren voor andere spelers. Een open research information infrastructuur beschermt ook de academische onafhankelijkheid en data soevereiniteit door de metadata en onderzoeksinformatie open te maken en te houden.