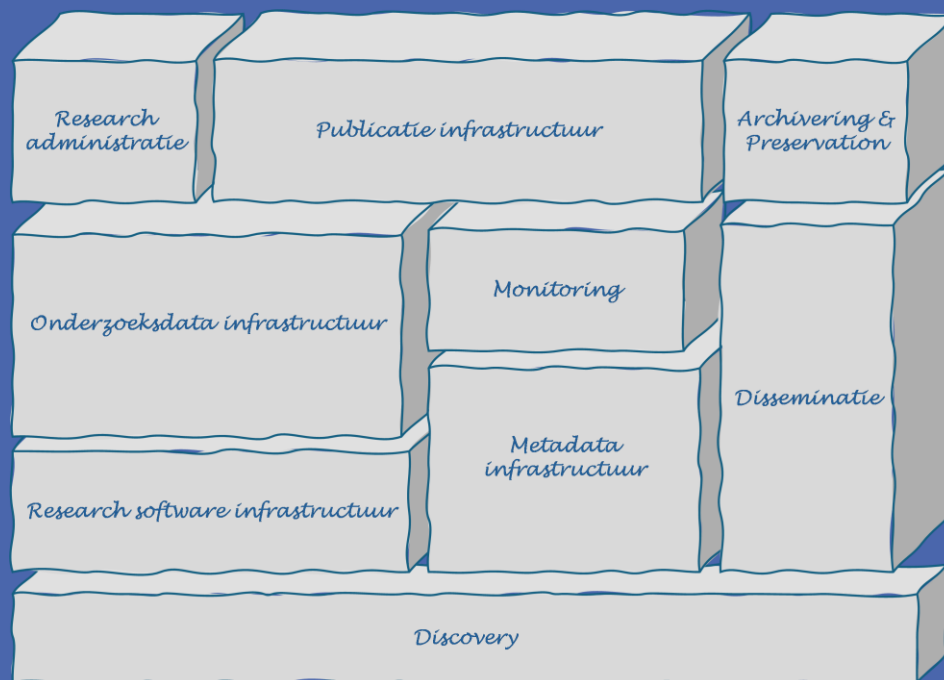


CONTOUREN OPEN SCIENCE INFRASTRUCTUUR

STRATEGISCH PLAN INTEGRALE INFRASTRUCTUUR VOOR OPEN SCIENCE DEEL 1



Projectgroep SPII: Till Bey; Patris van Boxel; Hubert Krekels; Jason Maassen; Marcel Ras;
Maurice Vanderfeesten; Claudia de Visser; Maurits van der Graaf; versie 1.0; 10-7-2025

Management samenvatting.....	4
1. Inleiding	8
1.1 Open Science infrastructuur.....	8
1.2 Scope van SPII Deel 1	11
1.3 Opzet project.....	12
2. Visie op Open Science Infrastructuur	13
2.1 Globale schets OSI	13
2.2 Visie op de Open Science Infrastructuur.....	13
2.3 Beoogde functies van OSI	13
2.3.1 Delen van FAIR onderzoeksresultaten	13
2.3.2 Functies van OSI als geheel.....	14
2.3.3 Randvoorwaarden	14
2.4 Vier kenmerken van de open science infrastructuur	15
2.5 Overige overwegingen	17
3. Contouren van de Open Science Infrastructuur	19
3.1 Infrastructurele context	19
3.2 Totstandkoming contouren OSI	20
3.3 Contouren van OSI	21
3.3.1 Research administratie en compliance met regelgeving.....	22
3.3.2 CRIS/repositories: optimalisatie- en efficiëntieslag en inzet als publicatiekanaal	24
3.3.3 Infrastructuur delen en hergebruik (sensitieve) data	28
3.3.4 Infrastructuur delen en hergebruik software	32
3.3.5 Open Research informatie, discovery en actieve disseminatie	33
3.3.6 Federatief afsprakenstelsel, aanvullende functionaliteit, ontwerpcriteria en randvoorwaarden	35
3.4 Naar een integrale referentiearchitectuur voor OSI	37
4. Discussie, conclusies en aanbevelingen voor SPII Deel 2.....	38
4.1 Discussie en evaluatie SPII Deel 1	38
4.2 Conclusie.....	40
4.3 Aanbevelingen voor SPII Deel 2	41

5. Referenties	45
Bijlage A: Methode – use-cases en federatieve schema’s.....	46
A1. Inleiding.....	46
A2. OSC: Use-cases en federatief schema	48
A3. RDSM: Use-cases en federatieve schema’s	58
A4. ORI: Use-cases en federatief schema	65
A5. Research administratie: Use-case en federatief schema	68
Bijlage B: Toelichting Scope SPII Deel 1	70
B1. Inleiding.....	70
B2. Open Science infrastructuur afgebakend van research infrastructuur:	70
B3. Federatieve infrastructuur als uitgangspunt	70
B4. RDSM, OSC en ORI.....	72
B5. Discipline-agnostisch in Deel 1.....	72
Bijlage C: Vertrekpunten voor de ontwikkeling van de OSI Referentiearchitectuur.....	74
Bijlage D: Suggesties voor de start van SPII Deel 2 en 3	79
Bijlage E: Deelnemers SPII Deel 1	81
Bijlage F: Glossary.....	83
Figuur 1 Proces stappen en presentatie in dit rapport	12
Figuur 2 Globale schets Open Science Infrastructuur	13
Figuur 3 OSI in context	19
Figuur 4 Meer gedetailleerde invullen van OSI	19
Figuur 5 Methode totstandkoming contouren OSI	20
Figuur 6 Contouren van OSI	21
Figuur 7 Overzicht voorstel werkprogramma SPII Deel 2	41
Figuur 8 Use-cases gekoppeld aan de onderzoekscyclus.....	46

DOELSTELLING

Het doel van het Strategisch Plan Integrale Infrastructuur Open Science (SPII) is het ontwikkelen van een plan voor een samenhangende, federatieve infrastructuur die onderzoekers en instellingen ondersteunt bij het faciliteren van Open Science. De aanleiding is het ontbreken van een gemeenschappelijk kader voor een Open Science Infrastructuur (OSI), hetgeen het risico van versnippering met zich meebrengt. SPII beoogt een gemeenschappelijk gedragen streefbeeld en streefarchitectuur te ontwikkelen dat een sturende en harmoniserende werking heeft zodat er stapsgewijs een adequate, landelijke Open Science Infrastructuur kan worden gerealiseerd.

In dit project is SPII Deel 1 ontwikkeld met een visie, een kader voor de infrastructuur en een fit-gap analyse t.o.v. de bestaande situatie. Wanneer SPII Deel 1 omarmd wordt, kan dit in SPII Deel 2 en 3 verder uitgewerkt worden.

SCOPE

De scope van SPII Deel 1 is kort samengevat:

- OSI (Open Science infrastructuur) is gepositioneerd als het Open Science-perspectief op de bredere digitale research infrastructuur.
- Een federatieve, modulaire opzet is gekozen als uitgangspunt om zoveel mogelijk aan te sluiten bij bestaande systemen.
- De focus van SPII Deel 1 ligt op Research Data & Software Management, Open Scholarly Communication en Open Research Information met nadruk op generieke functionaliteiten voor het FAIR delen van onderzoeksresultaten, relevant voor alle disciplines.

VISIE

De Open Science Infrastructuur (OSI) zorgt dat Nederlands onderzoek duurzaam, FAIR, en zo open als mogelijk gedeeld kan worden en dat de resultaten kunnen worden gearhiveerd, zichtbaar gemaakt en hergebruikt. Tevens aggregereert de infrastructuur onderzoeksinformatie vanuit verschillende componenten en maakt deze zo open als mogelijk toegankelijk. De OSI doet dat flexibel en pragmatisch door een federatieve, modulaire en integrale aanpak die openheid, digitale soevereiniteit en digitale duurzaamheid van de onderzoeksresultaten borgt. Daarbij worden vijf hoofdfuncties onderscheiden:

- Delen van FAIR onderzoeksresultaten
- Zichtbaarheid en disseminatie van de Nederlandse onderzoeksresultaten verhogen
- Monitoring van het gebruik d.m.v. tellers en metrics
- Open Research informatie uit de verschillende OSI-componenten wordt samengevoegd en beschikbaar gesteld.
- Archivering en long-term preservation van de Nederlandse onderzoeksresultaten.

ONTWIKKELMETHODE

Bij het ontwikkelen van de voorstellen voor OSI heeft de projectgroep op basis van de visie de federatieve opzet zoveel mogelijk vanuit het perspectief van de onderzoekers en onderzoeksondersteunende professionals ontwikkeld. Per domein zijn een aantal cruciale use-cases uitgewerkt (zie bijlage A), waarbij telkens gekeken is óf en zo ja, wat de noodzaak is van federatieve componenten. Op basis daarvan zijn federatieve schema's ontwikkeld, waarbij de gewenste federatieve oplossingen, de daarvoor noodzakelijke federatieve afspraken en de door de federatie te verbinden modules in kaart zijn gebracht. Op basis van deze federatieve schema's zijn vervolgens voorstellen ontwikkeld voor componenten van OSI.

BESCHRIJVING OSI COMPONENTEN

- **Research administratie en compliance met regelgeving**
 - **Situatie nu:** Onderzoekers worden geconfronteerd met toenemende administratieve verplichtingen (DMP, ethiek, AVG, licenties) bij de start van een onderzoeksproject. Er zijn momenteel meerdere instellingen bezig met het ontwikkelen van een systeem om de workflows van onderzoekers te ondersteunen en redundante invoer te vermijden.
 - **Doelstelling:** Inzetten op ketenintegratie door de voor Open Science relevante gegevens die in dit proces worden vastgelegd te harmoniseren en in te zetten bij de uitvoering en afronding van het onderzoeksproject. Dit door 'het meenemen' van relevante metadata en het automatiseren waar mogelijk.
 - **Voorgestelde oplossing:** Integratie van Open Science-gerelateerde metadata vanaf de start van het project (zoals DMP, ethiek toetsing) in het gehele onderzoeksproces.
- **Optimalisatie- en efficiëntieslag CRIS/repositories en inzet als publicatiekanaal:**
 - **Situatie nu:** Veel instellingen hebben een institutionele repository en een CRIS systeem ingericht ten behoeve van registratie en rapportage. E.e.a. betekent een forse inspanning in termen van mensen en middelen. De werkwijze verschilt sterk tussen de instellingen en er vindt veel dubbel werk plaats.
 - **Doelstellingen:**
 - Optimalisatie en efficiëntieverbetering van de huidige CRIS systemen en repositories.
 - Inzetten van deze op registratie en rapportage ingerichte systemen en inspanningen in het publicatieproces ten behoeve van (innovatieve) publicatievormen.
 - **Voorgestelde oplossingen:** De genoemde systemen koppelen met een federatieve oplossing en de inrichting van een nationale publicatieaggregator ten behoeve van niet-traditionele onderzoekspublicaties en van innovatieve publicatievormen zoals het PRC model. Deze nationale publicatieaggregator kan tevens dienen voor de archivering en longterm preservatie én een startpunt zijn voor beschikbaarstelling van wetenschappelijke informatie voor een breed publiek.
- **Infrastructuur delen en hergebruik (sensitieve) onderzoeksdata:**

- **Situatie nu:** Het landschap van RDM- infrastructurele componenten vertoont een grote diversiteit en is nog sterk in ontwikkeling. Wat betreft de Open Science aspecten is er geen of onvoldoende integratie van de keten research administratie → data management plan → data management omgeving → (long-term) datarepository en ontbreken geautomatiseerde data access mechanismen voor datasets met door condities beperkte toegang.
- **Doelstellingen:**
 - Ketenintegratie waar mogelijk door middel van federatie
 - Automatisering waar mogelijk om de onderzoekers te ontlasten
- **Voorgestelde oplossingen:**
 - Identificeren van de tijdens de verschillende stappen ontwikkelde metadata die relevant zijn voor het delen en hergebruik van de data en voor de reproduceerbaarheid van het onderzoek om deze tijdens het onderzoeksproces mee te nemen, te hergebruiken en verder aan te vullen.
 - Aanpak van automatisering van data access mechanismen.
- **Infrastructuur delen en hergebruik research software:**
 - **Situatie nu:** Met de implementatie van het Research Software Directory enkele jaren geleden zijn alle technische componenten aanwezig voor wat betreft het delen en hergebruik van research software. Het schort echter aan de adoptie ervan.
 - **Doelstellingen:** Bevorderen van de adoptie van delen en hergebruik research software door het ontwikkelen van breed gedragen Best Practices
 - **Voorgestelde oplossingen:** Ontwikkeling van Best Practices waar nodig.
- **Open research informatie, discovery en actieve disseminatie:**
 - **Situatie nu:** Een belangrijk deel van de research informatie over Nederlands onderzoek is niet open beschikbaar. Het deel dat wel open is, is gefragmenteerd beschikbaar. De discovery en disseminatie is gerealiseerd door het Netherlands Research portal dat nog lacunes kent qua functionaliteit.
 - **Doelstellingen:**
 - Een infrastructuur voor de beschikbaarstelling van open research informatie over Nederlands onderzoek met een zo compleet mogelijke dataverzameling die door alle partijen gebruikt en geanalyseerd kan worden.
 - Verbetering van de discovery en disseminatie van de Nederlandse onderzoeksresultaten met als resultaat een verhoogde (inter-)nationale zichtbaarheid.
 - **Voorgestelde oplossingen:**
 - De inrichting van een metadatahub, waar de informatie over onderzoek en de metrics bijeen wordt gebracht met als doel deze zoveel mogelijk open beschikbaar te stellen.
 - De inrichting van een discovery layer met dezelfde informatie als de metadatahub, op een wijze gepresenteerd die geschikt is voor discovery en actieve disseminatie naar diverse (inter-)nationale zoekmachines, aggregatie-platformen e.d.

DISCUSSIE SPII DEEL 1, CONCLUSIES EN AANBEVELINGEN VOOR SPII DEEL 2

Consultatie stakeholders: Er heeft een beperkte consultatie plaatsgevonden, waaruit een behoefte aan bredere betrokkenheid in een vervolg naar voren kwam. Daarnaast werden als belangrijkste feedbackpunten genoemd:

- De uitgewerkte use-cases dekken vooral de start en afronding van onderzoek en niet of nauwelijks de uitvoeringsfase.
- Sommige infrastructurele voorstellen (zoals een nationale publicatieaggregator) zijn nog onvoldoende besproken met stakeholders.
- Sommigen bepleiten een meer abstracte, functionele benadering van OSI, los van bestaande systemen.

Conclusie: Uit de reacties van de consultatieronde blijkt dat er veel belangstelling is voor een visie en strategie op het gebied van Open Science infrastructuur. Dit rapport kan gezien worden als een eerste stap en een aanleiding voor een verdere ontwikkeling van OSI.

Aanbevelingen voor SPII Deel 2:

- We bevelen aan om SPII Deel 2 op te knippen in Deel 2 en Deel 3.
- SPII 2 richt zich op de technische uitwerking van SPII Deel 1 met als voorgestelde deliverables:
 - Technische infrastructurele elementen met o.a. een referentiearchitectuur en een metadataschema en technische standaarden en afspraken over koppelvlakken
 - Voorstel voor een federatief afsprakenstelsel
 - Roadmap voor OSI met ontwikkeling- en implementatievoorstellen voor de periode 2026-2029.

Wanneer de deliverables van SPII Deel 2 omarmd worden, kunnen in SPII Deel 3 de organisatorische aspecten ontwikkeld worden in nauwe afstemming met ontwikkelingen rond EOSC-NL.

1. INLEIDING

1.1 OPEN SCIENCE INFRASTRUCTUUR

INTRODUCTIE

Een adequate nationale Open Science infrastructuur (OSI) maakt het voor onderzoekers gemakkelijker en aantrekkelijker om Open Science in hun onderzoekspraktijk toe te passen. Dergelijke infrastructuur helpt bij het voldoen aan Open Science principes (bijv. FAIR research data en software) en de gedragscode wetenschappelijke integriteit (bijv. navolgbaarheid en transparantie)¹. Daarnaast kan een adequate OSI onderzoekers deels ondersteunen bij het voldoen aan wet- en regelgeving (zoals de AVG) en instellings- en instituutsbeleid (t.a.v. veiligheid, datamanagement en ethiek).

De uitgangspunten voor een dergelijke infrastructuur voor Open Science zijn al in een vroeger stadium geformuleerd². De ontwikkeling van Open Science infrastructuur is een belangrijk onderdeel van het *NPOS 2030 Ambition Document and Rolling Agenda*³ en de UNL Open Science Agenda^{4,5}. Het NPOS Ambition document bevat onder de kop *'Make Open Science possible through Open Infrastructures'* een aantal doelstellingen, waarnaar in dit rapport verwezen wordt.

DEFINITIE OPEN SCIENCE INFRASTRUCTUUR

De projectgroep heeft als definitie voor Open Science infrastructuur de definitie van OSNL gehanteerd:

OSI betreft tools, software, workflows, platforms en digitale services die nodig zijn voor de implementatie van Open Science. OSI faciliteert onderzoekers bij het management en delen van research data, software, publicaties, hardware en andere research output, zoals tools, protocollen en methoden. OSI helpt onderzoekers om Open Science in de praktijk te brengen door het mogelijk maken van open access publiceren en het delen van data en software. Daarnaast kan OSI bijdragen aan de toegankelijkheid, vindbaarheid, en hergebruik van research output door burgers en andere partijen buiten academia, bijvoorbeeld door de maatschappelijke impact van wetenschappelijke kennis te verhogen. Tenslotte kan OSI de transparantie en reproduceerbaarheid van wetenschappelijk onderzoek verhogen, hetgeen bijdraagt aan de kwaliteit van het onderzoek en het vertrouwen in de wetenschap^a.

HUIDIGE SITUATIE

De belangrijkste ontwikkelingen zijn:

^a Vertaald uit Engels. Zie Call for proposals Open Science Infrastructure, 2024, 1st OSNL

- Veel universiteiten en hogescholen hebben reeds onderdelen van een Open Science infrastructuur geïmplementeerd of in ontwikkeling.
- SURF heeft een innovatiezone ‘Open Science versterken’ waarin ook infrastructuur wordt versterkt of ontwikkeld⁶.
- Een recente call van Open Science NL op dit gebied leverde veel projectvoorstellen op, hetgeen een grote behoefte laat zien⁷.
- Er is een pilotproject opgestart voor een Nederlandse EOSC node.

KADER VOOR OPEN SCIENCE INFRASTRUCTUUR GEWENST

Uit het bovenstaande blijkt dat er veel initiatieven zijn ten aanzien van infrastructuur voor Open Science, waarbij een gezamenlijk beeld van de infrastructuur voor Open Science behulpzaam zal zijn. Afstemming, coördinatie en prioritering van bestaande infrastructuur en van ontwikkelingen van nieuwe infrastructuur verloopt echter moeizaam vanwege het ontbreken van een algemeen, breed gedragen kader voor Open Science infrastructuur. Zo’n gemeenschappelijke aanpak van OSI is mede van belang uit oogpunt van kostenbeheersing en doelmatige besteding van de middelen voor de betrokken onderzoeksinstituten. Een gemeenschappelijk en breed gedragen kader, ook wel metaplan genoemd, dient gebaseerd te zijn op een algemene, breed gedragen visie op de voor universiteiten en hogescholen en overige kennisinstellingen benodigde integrale infrastructuur voor Open Science. Daarin kunnen dan de infrastructuurcomponenten worden geïdentificeerd en de eigenschappen worden omschreven waaraan deze idealiter zouden moeten voldoen^{1,6}. Dit idee heeft geleid tot een door de betrokken stakeholders gedragen project onder de naam *Strategisch Plan Integrale Infrastructuur voor Open Science (SPII)*. Dit rapport beschrijft de resultaten van SPII Deel 1.

SPII IN TWEE DELEN

De doelstellingen voor het finale SPII zijn als volgt:

- **Biedt een streefbeeld van de integrale OSI met de gewenste modulaire componenten en onderlinge connecties:** Het strategische plan bevat een overzicht van gewenste modulaire componenten van de Open Science infrastructuur met de eigenschappen waaraan deze idealiter zouden moeten voldoen en wat de ‘gap’ is t.o.v. de reeds beschikbare infrastructuur d.m.v. fit/gap analyses. Tevens omvat het plan een beschrijving van de gewenste onderlinge connecties.
- **De gesignaleerde *gaps* met de huidige infrastructuur fungeren als richtsnoer bij vervanging en/of (door-)ontwikkeling** van bestaande of nieuwe infrastructuurcomponenten en heeft daardoor een sturende en harmoniserende werking zodat stapsgewijs een adequate Open Science infrastructuur wordt bereikt.
- **Wordt periodiek bijgesteld:** Met de voortgaande technische ontwikkelingen en veranderingen in de beschikbaarheid van lokale, nationale en internationale componenten zal de Open Science infrastructuur naar verwachting periodiek bijgesteld moeten worden. Gelet op de doorlooptijd van de ontwikkelplannen is het streven dat het in 2025 op te leveren SPII voor Open Science richtinggevend kan zijn voor ontwikkeltrajecten in de periode 2026-2030 en dat halverwege deze periode een update nodig is voor de periode daarna.

De doelstelling van SPII Deel 1 is om een visie op OSI te formuleren die breed gedragen zal worden én een eerste blauwdruk op te leveren die de verschillende actoren overtuigt van de waarde van een dergelijk strategisch plan, waardoor er draagvlak ontstaat voor een verdere uitwerking (voorzien in SPII Deel 2).

1.2 SCOPE VAN SPII DEEL 1

1. Open Science infrastructuur afgebakend van research infrastructuur:

- Het is de bedoeling om een infrastructuurplan te ontwikkelen voor Open Science. Daarbij is het onderscheid met de digitale research infrastructuur (zoals bijvoorbeeld High Performance Computing, GWI's etc.) belangrijk. Anders gezegd: OSI is het Open Science perspectief op de (digitale) research infrastructuur. De aansluiting op meer op disciplines gerichte Open Science infrastructurele componenten en op disciplinaire research infrastructuren komt in Deel 2 aan bod.

2. Federatieve infrastructuur als uitgangspunt:

- Het uitgangspunt is SPII zoveel mogelijk te baseren op wat er al is aan bestaande systemen. Daarom stelt SPII een federatieve/federatie-achtige opzet voor OSI voor, modulair opgebouwd en integraal van opzet. Een te ontwikkelen referentiearchitectuur voor OSI (à la HOSA) maakt deel uit van deze opzet.
- SPII Deel 1 focust op de technische vormgeving. De voor een federatieve structuur benodigde governance, organisatie, middelen en commitment vormen deel van het voorgestelde SPII Deel 3.

3. Research Data & Software Management (RDSM), Open Scholarly Communication infrastructuur (OSC) en Open Research Information (ORI) infrastructuur:

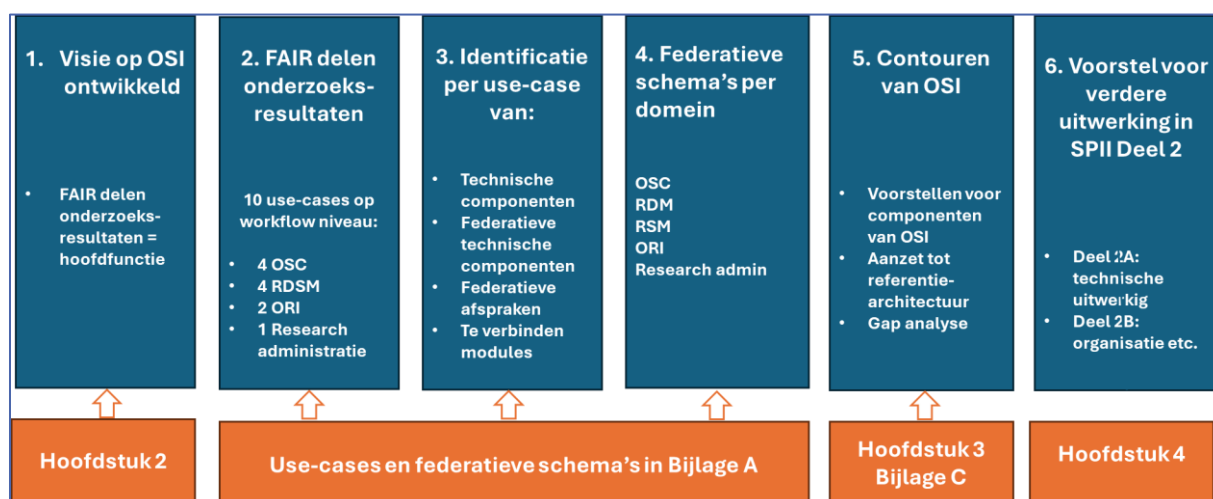
- SPII focust op een OSI infrastructuur voor Research Data & Software Management, Open Scholarly Communication infrastructuur en Open Research Information en op het aspect FAIR delen van onderzoeksresultaten daarbinnen. SPII verwacht daarmee een belangrijk deel van de behoefte aan OSI te dekken.

4. Discipline-agnostisch in Deel 1:

- SPII Deel 1 richt zich op OSI-functionaliteiten die relevant zijn voor Nederlandse onderzoeksprojecten ongeacht hun discipline. De focus ligt daarbij in eerste instantie op gegevens die de instellingen kunnen en willen delen op een effectieve en efficiënte manier door middel van een federatieve infrastructuur. De aansluiting op meer op disciplines gerichte Open Science infrastructurele componenten en op disciplinaire research infrastructuren komt in Deel 2 aan bod.

De scope is uitgebreider toegelicht in Bijlage B.

1.3 OPZET PROJECT



Figuur 1 Proces stappen en presentatie in dit rapport

Dit plan is ontwikkeld door de projectgroep SPII Deel 1 – een kleine groep experts, die vrijgemaakt waren door hun instellingen voor een beperkte tijd (100 uur) en een beperkte periode (maart-juli 2025). Daarnaast bestond de projectgroep uit een projectleider en een informatiearchitect, die hiervoor enkele dagen per week beschikbaar hadden in de genoemde periode.

In een eerste stap heeft de projectgroep een visie op OSI ontwikkeld. Deze wordt gepresenteerd in hoofdstuk 2. In deze visie werd als hoofdfunctie van OSI het FAIR delen van onderzoeksresultaten geïdentificeerd.

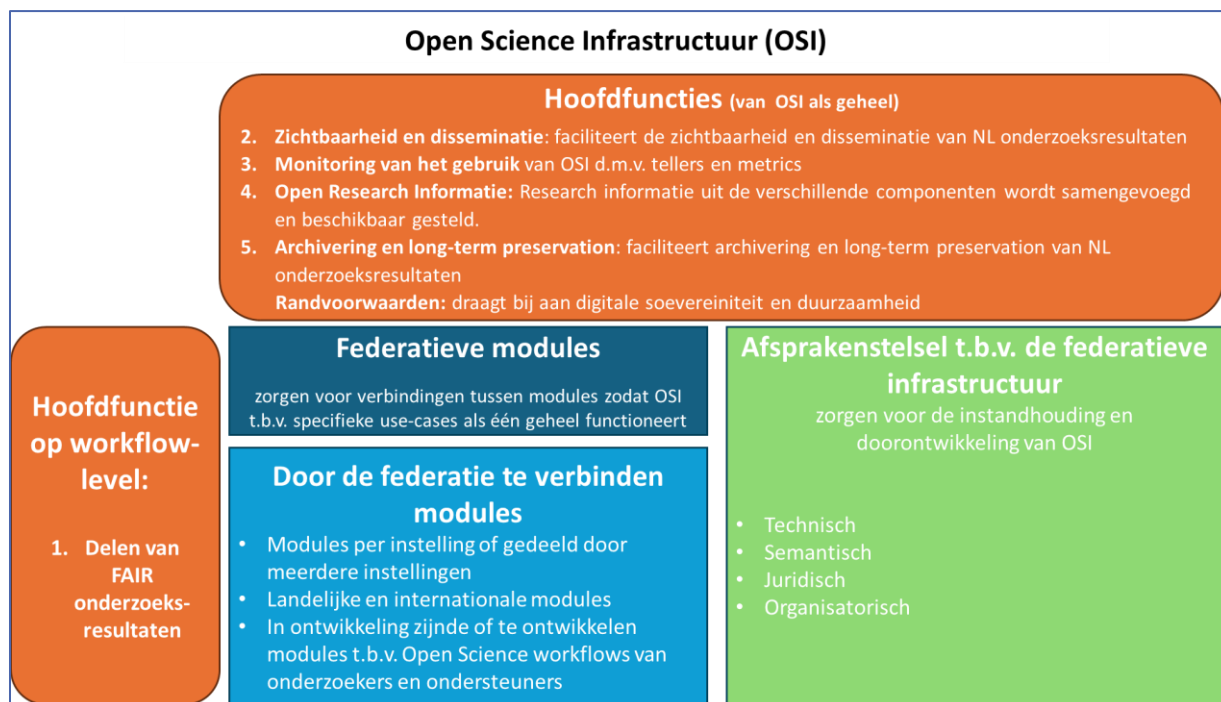
In een tweede stap werd gewerkt volgens een methode (use-cases → federatieve schema's → voorstellen voor de OSI infrastructuur; zie voor een meer gedetailleerde beschrijving 3.2).

Begin juni zijn de eerste resultaten beschreven in versie 0.7, die vervolgens is besproken in een consultatieronde met o.m. een aantal op de fit/gap-analyse gerichte sessies. Vanwege de tijdsdruk zijn hiervoor een beperkt aantal stakeholders geraadpleegd (zie voor alle deelnemers Bijlage E). Dit heeft geleid tot fit/gap analyses per domein. In de praktijk focusten deze sessies voornamelijk op de gaps, vandaar dat de resultaten ervan weergegeven zijn als gap analyses in hoofdstuk 3.

Overige ontvangen feedback van de geraadpleegde stakeholders is óf verwerkt in de voorliggende versie óf weergegeven in de discussie (zie hoofdstuk 4) óf weergegeven in Bijlage D met de suggestie om deze waar mogelijk in Deel 2 op te pakken.

2. VISIE OP OPEN SCIENCE INFRASTRUCTUUR

2.1 GLOBALE SCHETS OSI



Figuur 2 Globale schets Open Science Infrastructuur (blauw en groen) en 5 beoogde hoofdfuncties (oranje)

In figuur 2 wordt een globale schets van de Open Science infrastructuur gepresenteerd, die hieronder verder wordt toegelicht.

2.2 VISIE OP DE OPEN SCIENCE INFRASTRUCTUUR

De Open Science Infrastructuur zorgt dat Nederlands onderzoek duurzaam en zo open als mogelijk gedeeld kan worden en dat de resultaten kunnen worden hergebruikt, zichtbaar gemaakt, en gearchiveerd. Tevens brengt het systeem onderzoeksinformatie vanuit de verschillende componenten samen en stelt deze zo open als mogelijk beschikbaar. OSI doet dat flexibel en pragmatisch door een federatieve, modulaire en integrale aanpak die openheid en digitale soevereiniteit van de kennisinstellingen borgt.

2.3 BEOOGDE FUNCTIES VAN OSI

2.3.1 DELEN VAN FAIR ONDERZOEKSRESULTATEN

OSI faciliteert het delen van FAIR onderzoeksresultaten van Nederlandse kennisinstellingen. Het betreft het FAIR delen van onderzoeksresultaten op het niveau van workflows van onderzoekers en onderzoeksondersteunende professionals.

Met onderzoeksresultaten wordt onder meer bedoeld:

- Publicaties, bijvoorbeeld preprints, wetenschappelijke artikelen, vak- en wetenschappelijke publicaties, conferentieoutput, boeken en boekhoofdstukken, dissertaties, monografieën.
- Software, bijvoorbeeld computermodellen, algoritmes, softwarecodes, beschrijvingen, data en software verpakt in *executables*.
- Data(sets), bijvoorbeeld (beschrijvingen van) onderzoeksdata, geanalyseerde data, databanken, clinical trial data, synthetische data, verzamelde data
- Modellen, bijvoorbeeld prototypen, gebruiksvoorwerpen, (door)ontwikkelde apparatuur voor onderzoek of als resultaat van onderzoek.
- Protocollen (onderzoeksprotocollen, interventieprotocollen, instrumentatieprotocollen), logboeken, lab journals.

Het FAIR delen van onderzoeksresultaten dient ondersteund te worden door een afsprakenstelsel:

- **Findable:** gemeenschappelijke PIDs (Persistent Identifiers), metadatastandaarden
- **Accessible:** d.m.v. gemeenschappelijke afspraken over onder meer het toegangs- en rechtenbeheer volgens het adagium ‘Open waar mogelijk, gesloten waar nodig’, gebruik van gemeenschappelijke licenties e.d.
- **Interoperable:** gemeenschappelijke metadata elementen; gemeenschappelijke (vaak domeinspecifieke) vocabulaires/ontologieën
- **Reusable:** gemeenschappelijke licenties, gemeenschappelijke templates voor contextuele documentatie, harmoniseren van definities, zoals bijvoorbeeld over sensitiviteit van datasets.

2.3.2 FUNCTIES VAN OSI ALS GEHEEL

Zichtbaarheid en disseminatie van Nederlandse onderzoeksresultaten: Dit vereist onder andere rijke metadata en de verspreiding van deze metadata naar (inter-)nationale aggregatieplatformen, zoekmachines en databases/repositories.

Monitoring van het gebruik: Dit gebeurt door middel van tellers en metrics, die het gebruik (en de reviews) van de verschillende onderdelen bijhouden.

Open Research informatie: Research informatie uit de verschillende componenten van het systeem wordt samengevoegd en open beschikbaar gesteld. Gebruikers kunnen deze informatie naar zich toe halen en combineren met andere datasets en analyseren. Dit laatste wordt niet gezien als onderdeel van OSI.

Archivering en long-term preservation: Deze functie beoogt de archivering en long-term preservation van de onderzoeksresultaten. De archivering en preservation draagt bij aan de digitale soevereiniteit.

2.3.3 RANDVOORWAARDEN

- **Digitale soevereiniteit:** OSI beoogt bij te dragen aan de digitale soevereiniteit van de kennisinstellingen door (stapsgewijs) de academische kernwaarden te waarborgen, inclusief data-autonomie en het minimaliseren van vendor lock-ins. We stellen voor een pragmatische
-

benadering te volgen t.a.v. commerciële aanbieders. Bij inzet van componenten van commerciële aanbieders in de OSI gelden bij het inkoopproces criteria t.a.v. publieke waarden om vendor-lock-in te voorkomen en data-autonomie te borgen⁸, het gebruik van open standaarden, én een exit strategie, zodat de provider vervangbaar is (*substitutable providers*⁹) en de instellingen zonder verlies van data later kunnen overschakelen naar andere (niet-commerciële) providers en (open-source) componenten.

- **Duurzaamheid:** OSI dient duurzaam opgezet te worden. Dit betekent o.m. dat er aan een aantal financiële randvoorwaarden voldaan moet worden en een gebalanceerd financieringsmodel uitgewerkt dient te worden. Het gaat o.m. om zaken als projectfinanciering uitsluitend voor projecten, een structurele nationale financiering voor core infrastructuur componenten zoals een interoperabiliteitslaag en authenticatiediensten met eventueel een verdienmodel gebaseerd op diensten, een reserve voor de operationele kosten van ca. 1 jaar en eigen financiering door instellingen van bijvoorbeeld lokale storage oplossingen en instellingspecifieke infrastructuurcomponenten¹⁰. E.e.a. wordt verder uitgewerkt in SPII Deel 2. Een ander aspect van duurzaam betreft de exit-strategie, waarbij gegarandeerd dient te worden dat de data duurzaam elders bewaard en toegankelijk kan worden gemaakt en dat het verzamelen van die data op een andere wijze kan worden voortgezet.
- **Compliant:** OSI moet gebruikers – naast Open Science principes - in staat stellen te voldoen aan wet- en regelgeving (e.g., AVG), aan instellings- en instituutsbeleid (zoals regels t.a.v. research data, ethiek en veiligheid), en aan de gedragscode wetenschappelijke integriteit (transparantie, navolgbaarheid).
- **Dataveiligheid:** Dataveiligheid dient in OSI gewaarborgd te zijn door bescherming van datasets waar dat nodig is, door te zorgen dat de integriteit van dat geborgd is en dat er beveiliging is tegen cybercriminaliteit.
- **Kosteneffectief:** Samenwerking aan en integratie waar mogelijk van een gemeenschappelijke OSI leidt tot kostenbeheersing en doelmatige besteding van middelen. Samenwerking is ook noodzaak: individuele kennisinstellingen beschikken naar verwachting niet over voldoende middelen om zelfstandig een OSI-infrastructuur te ontwikkelen en te onderhouden.
- **Aanvullend, aansluiten en interoperabel:** OSI is aanvullend, focust op Open Science en sluit aan op bestaande, ontwikkelde of in ontwikkeling zijnde infrastructuren (zowel lokaal als nationaal en internationaal), zoals federatieve authenticatie (SURFconext), federatieve discovery via Netherlands Research Portal, een aantal bij OSNL ingediende projectvoorstellen voor gemeenschappelijke infrastructuren, de in ontwikkeling zijnde DCC/Tier2 infrastructuur, op *Large Scale Research infrastructures* (bijv. GWI's, ERIC's), en op internationale infrastructuren zoals ORE, OpenAIRE en EOSC.

2.4 VIER KENMERKEN VAN DE OPEN SCIENCE INFRASTRUCTUUR

De OSI bestaat uit: (a) instellingspecifieke componenten; (b) componenten die met meerdere instellingen gedeeld worden; (c) landelijke componenten; (d) internationale componenten en (e) is gebaseerd op (inter-)nationale standaarden en afspraken over interoperabiliteit op vier niveaus: technisch, semantisch, juridisch en organisatorisch.

FEDERATIEF

- **Federatie door gemeenschappelijke architecturale afspraken:** De componenten van de infrastructuur zijn verbonden via structurele, flexibele, gemeenschappelijke architecturale afspraken met het oog op gemeenschappelijke functionaliteit door interoperabiliteit.
- **Federatie op landelijk niveau:** Voor het Strategische Plan Integrale Infrastructuur Open Science (SPII) wordt de federatieve structuur ontwikkeld op supra-institutioneel, landelijk niveau.
- **Aansluiting op internationale infrastructuren:** De federatieve opzet zorgt voor een samenhangende en functionele technische infrastructuur op landelijk niveau, die door middel van koppelvlakken aangesloten kan worden op internationale infrastructuren zoals OpenAIRE, Open Research Europe en EOSC.
- **Federatieve afspraken:**
 - Gemeenschappelijke afspraken over standaarden en software voor technische integratie.
 - De federatieve infrastructuur is tevens gebaseerd op organisationele aspecten zoals een stelsel van afspraken tussen de instellingen over regels over deelname en toegang, kwaliteitsbewaking van de gegevens, verdeling van de financiële kosten en governance^{11,12}.
- **Behoud van autonomie:** Het federatieve model behoudt de autonomie van de infrastructurele componenten die beheerd worden door de deelnemende instellingen en bevordert zo het eigenaarschap en verantwoordelijkheid daarvoor. Vereist is wel dat de architecturale en federatieve afspraken blijvend worden nageleefd.

MODULAIR

- **Modules op meerdere niveaus:** De modules van de Open Science infrastructuur kunnen modules van één of meerdere instellingen zijn of zich op landelijk of internationaal niveau bevinden.
 - a) Met instellingspecifieke modules als basis blijven de instellingen in controle over de belangrijkste data waarover elke instelling beschikt – de ‘data-kroonjuwelen,’ zoals o.m. gegevens over onderzoekers en onderzoeksdata. Deze brondata worden zoveel mogelijk losgekoppeld van de applicatielaag⁸. Hierdoor behouden de instellingen zoveel mogelijk data-autonomie en ruimte voor specifiek instellingsbeleid. Voorbeelden zijn institutionele CRIS-systemen en institutionele repositories.
 - b) Met componenten die met meerdere instellingen gedeeld worden kunnen gelijksoortige instellingen modules beheren en ontwikkelen, en zo efficiëntievoordelen behalen. Voorbeelden van componenten, die met meerdere instellingen gedeeld worden, zijn een Data Access Broker dienst of PubliNova.
 - c) De landelijke modules en de gemeenschappelijke aansluitingen op internationale infrastructuren bieden eveneens efficiëntievoordelen en/of inzicht in informatie over de onderzoeksresultaten. Het Netherlands Research portal is een voorbeeld.
- **Focus op verbindingen tussen modules:** De focus van het Strategische Plan Integrale Infrastructuur Open Science (SPII) ligt op de federatieve koppelingen tussen deze modules en op de gemeenschappelijke afspraken, en niet zozeer op de modules zelf.

INTEGRAAL

- **Ondersteuning Research Life cycle:** De Open Science infrastructuur is integraal omdat de infrastructuur meerdere fasen van de research life cycle beoogt te ondersteunen. Dit betekent dat de infrastructuur ontworpen is om ondersteuning te bieden bij Research Data en Software

Management (RDSM), Open Scholarly Communication (OSC) en Open Research Information (ORI)^b.

- **Integratie is kostenefficiënt:** Sommige componenten zullen specifiek zijn voor een van deze domeinen, andere componenten voor meerdere. Deze integratie maakt de opzet, ontwikkeling en beheer kostenefficiënter en het gebruik ervan veiliger.

OPEN

- **Open infrastructuur:** De infrastructuur is open waar mogelijk, en gesloten waar nodig. Dit betekent dat sommige (delen van) componenten voor een beperkte groep toegankelijk zullen zijn vanwege beheer- en/of veiligheidsoverwegingen.
- **Open verbindingen:** De OSI is verbonden met open standaarden en/of API's en volgt de internationale ontwikkelingen om open standaarden te definiëren, uit te breiden of te moderniseren.
- **Open-source:** Voor de componenten van de OSI wordt waar mogelijk gebruikgemaakt van open-source software, van open-source hardware en van aanbieders die data-autonomie respecteren en waarbij er geen sprake is van 'vendor lock-in'. Het streven is om ook bij te dragen aan bestaande projecten van open-source communities en door eventuele eigen softwareontwikkelingen open-source te maken.

2.5 OVERIGE OVERWEGINGEN

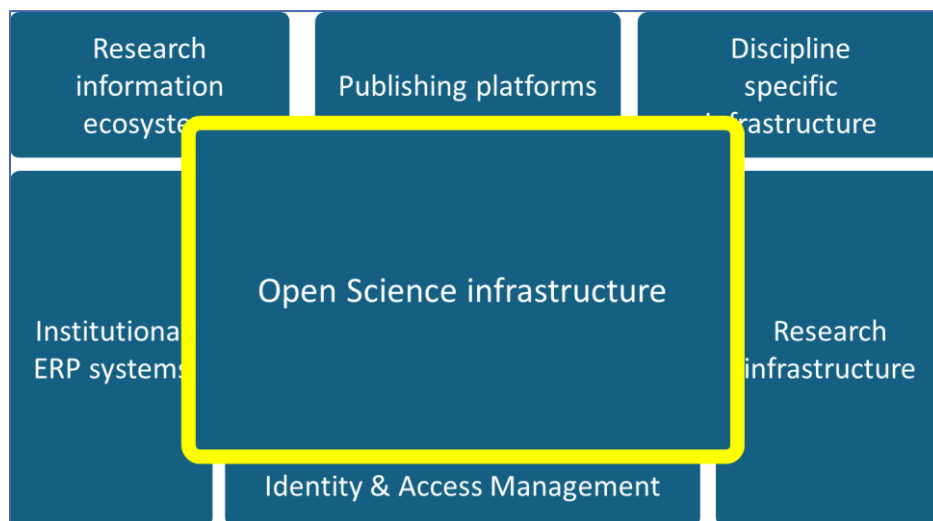
- **Geén dichtgetimmerd plan, maar federatieve oplossingen:** SPII is geen dichtgetimmerd plan, waarin precies wordt voorgeschreven wat er moet worden uitgevoerd. Dergelijke plannen worden in de praktijk zelden uitgevoerd. SPII wil daarentegen federatieve oplossingen voorstellen om bepaalde functies te realiseren: dit zorgt ervoor dat bij een eventuele implementatie de creativiteit van de ontwikkelaars én de nieuwste technische ontwikkelingen optimaal benut worden.
- **Meerwaarde:** De meerwaarde van de Open Science infrastructuur ligt vanzelfsprekend in het vergemakkelijken en bevorderen van Open Science voor het onderzoek aan de Nederlandse kennisinstellingen. Meer specifiek voor de OSI zelf betekent dit dat de inrichting ervan wordt gegroepeerd rond een aantal functies die vanwege hun instellingsoverstijgend belang tegelijkertijd de businesscase ervoor vormen. De meerwaarde van een gezamenlijke infrastructuur voor deze functies ligt in de kostenefficiëntie voor de deelnemende instellingen en het bevorderen van de samenwerking van onderzoekers tussen de deelnemende instellingen.
- **Flexibiliteit in de implementatie:** De implementatie van de OSI zal voor de verschillende functies én voor de verschillende deelnemers verschillende tijdspaden kunnen volgen. Daarbij wordt in de opzet ook rekening gehouden met verschillende niveaus van deelname. Verder dient de opzet van de infrastructuur zodanig vormgegeven te zijn dat er flexibel kan ingespeeld worden op nieuwe (technische) ontwikkelingen.

^b Dit komt overeen met twee speerpunten van de NPOS Rolling Agenda: 'Towards open scholarly communication' en 'towards FAIR and open research outputs'.³

- **Afsprakenstelsel, governance en financiering:** Een federatief, technisch systeem kan alleen functioneren wanneer er een duidelijk afsprakenstelsel aan ten grondslag ligt. In deze opzet wordt gekozen om eerst een ontwerp voor de gewenste technische en semantische infrastructuur te ontwikkelen en in een vervolgstap de daarbij behorende afspraken, governance- en financieringsmodel te ontwikkelen die zoveel mogelijk aansluiten op bestaande structuren en modellen.
- **Flexibiliteit m.b.t. standaarden:** vaak is het niet per se noodzakelijk voor een federatieve connectie dat de twee systemen bijvoorbeeld dezelfde metadata standaard hanteren. Het is mogelijk om een vertaling (crosswalk) in de API.

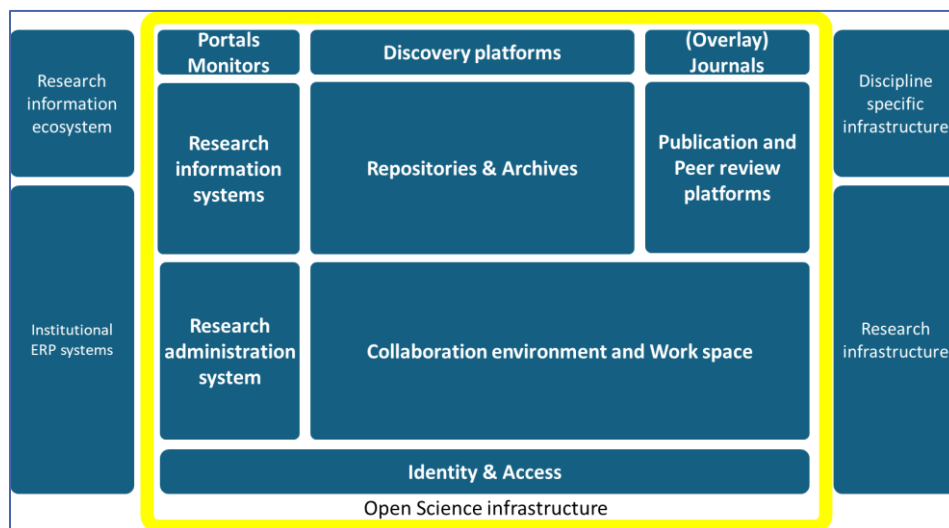
3. CONTOUREN VAN DE OPEN SCIENCE INFRASTRUCTUUR

3.1 INFRASTRUCTURELE CONTEXT



Figuur 3 OSI in context

OSI staat niet op zichzelf, maar is ingebed in het bredere landschap van bestaande infrastructuur die onderzoek ondersteunt. Het gaat daarbij onder meer om veelal discipline-gerichte research infrastructuur, maar ook om het ecosysteem van uitgevers, aanbieders van metadata et cetera. In figuur 3 is deze context geschetst.

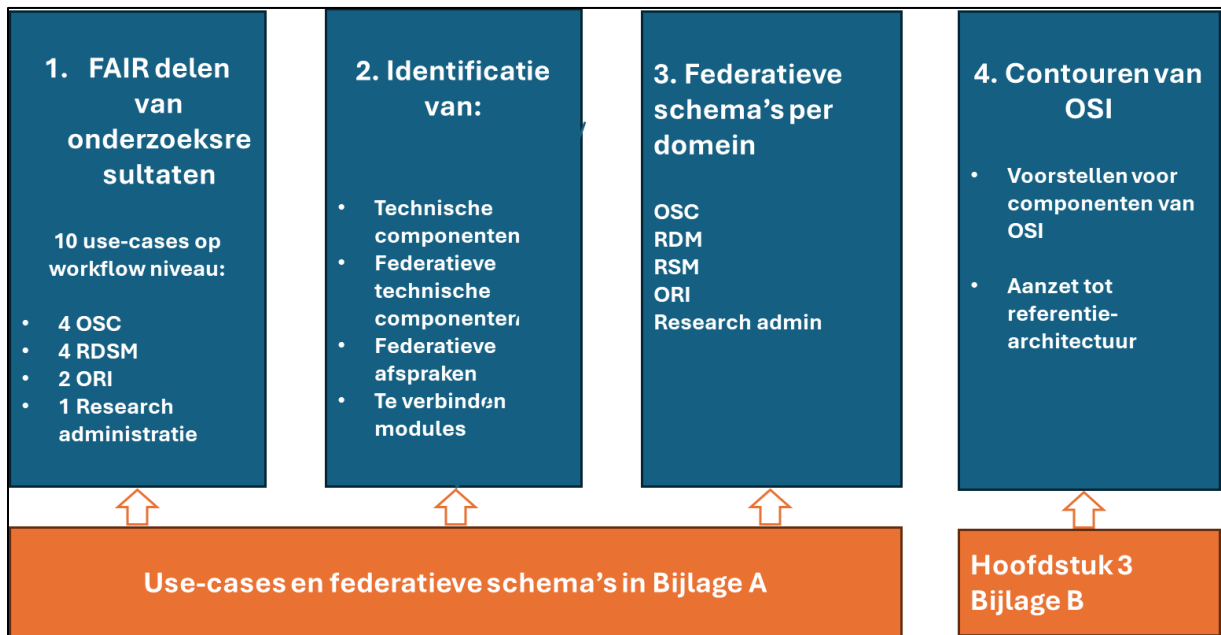


Figuur 4 Meer gedetailleerde invullen van OSI

Inzoomend op OSI behoren de componenten ervan tot een aantal categorieën die functies op hoog niveau ondersteunen, meestal in combinatie (zie figuur 4). Enkele voorbeelden:

- Zichtbaarheid wordt gefaciliteerd door portals en monitors die informatie leveren over het Nederlandse onderzoek met o.m. discovery platforms met zoekfunctionaliteit en overlay journals die gecureerde collecties van publicaties bundelen.
- Geïntegreerde researchadministratie-systemen en onderzoeksinformatie-systemen (CRIS-systemen) werken samen om monitoring en rapportage mogelijk te maken en compliance met regelgeving te waarborgen.

3.2 TOTSTANDKOMING CONTOUREN OSI



Figuur 5 Methode totstandkoming contouren OSI

Figuur 5 schetst de gevolgde methode voor de ontwikkeling van de contouren van OSI.

De projectgroep heeft 10 use-cases van workflows van onderzoekers en onderzoeksondersteunende professionals opgesteld. Een groot deel van deze use-cases is gericht op het FAIR delen van onderzoeksresultaten, de hoofdfunctie van OSI. Dit met het oog op het zoveel mogelijk bottom-up ontwikkelen van OSI vanuit het perspectief van de onderzoekers en onderzoeksondersteunende professionals.

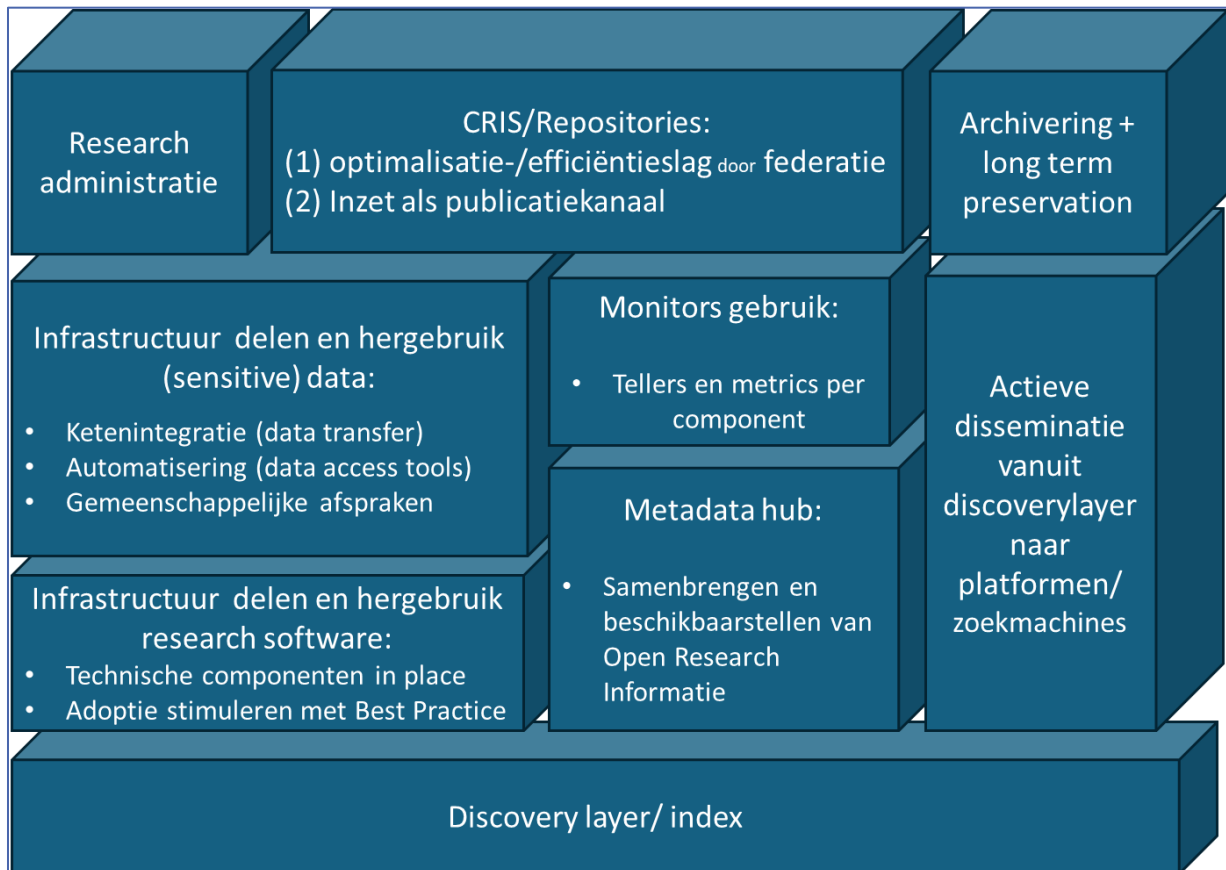
Samengevat omvat de methode de volgende stappen:

1. Per use-case werd op basis van de stappen in de workflow de benodigde technische elementen geïdentificeerd en welke daarvan federatief zouden moeten zijn, de sociale en menselijke elementen die een rol spelen bij de stappen in de workflow en de federatieve afspraken die nodig zijn wat betreft standaarden en interoperabiliteit.
2. Per domein zijn de in de use-cases geïdentificeerde federatieve oplossingen in een federatief schema weergegeven met daarin de lijst van federatieve afspraken (technisch, semantisch, juridisch) en de door de federatieve oplossing te verbinden modules.

3. De federatieve oplossingen zoals gepresenteerd in de federatieve schema's vormden de basis voor de voorstellen in de volgende paragraaf voor de verschillende componenten van de OSI infrastructuur.

De use-cases en de federatieve schema's zijn beschreven in Bijlage A.

3.3 CONTOUREN VAN OSI



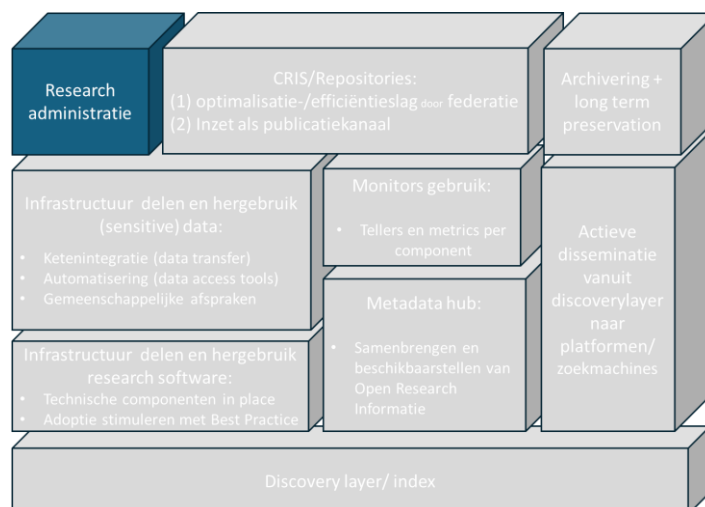
Figuur 6 Contouren van OSI

In figuur 6 zijn de contouren van OSI weergegeven in een blokkenschema. De beschrijving van de verschillende samenhangende blokken volgt in de hierna volgende paragrafen.

3.3.1 RESEARCH ADMINISTRATIE EN COMPLIANCE MET REGELGEVING

HUDIGE SITUATIE

- De administratieve component van wetenschappelijk onderzoek is de afgelopen jaren sterk in omvang en complexiteit toegenomen. Onderzoekers worden geconfronteerd met een groeiend aantal verplichtingen vanuit onderzoeksfinanciers en instellingen op het gebied van research data management, ethische toetsing, privacywetgeving (zoals de AVG), datalicenties en contractuele voorwaarden.



Voorafgaand aan onderzoek zullen onderzoekers vaak een subsidieaanvraag doen. Daarbij zijn allerlei document-uitwisselingen en communicatie-workflows nodig tussen onderzoeker, project control, onderzoeksdirecteur, data steward, juridische afdeling, etc. In sommige gevallen wordt in dit stadium ook al een DMP (datamanagementplan) opgesteld, gecontroleerd en goedgekeurd. Deze ontwikkelingen maken dat research administratie en compliance met regelgeving een integraal onderdeel vormen van het onderzoeksproces. Wat betreft Open Science bepalen deze gegevens bij de start van een onderzoeksproject in hoge mate in hoeverre het onderzoeksproces en de onderzoeksresultaten open, open onder condities dan wel gesloten dient plaats te vinden. Daarom ziet de projectgroep dit als een belangrijk element van OSI.

- Volgens onze informatie zijn er meerdere instellingen (soms universiteiten of hogescholen, soms faculteiten binnen een universiteit of hogeschool) bezig met het ontwikkelen van een systeem om in eerste instantie de workflows van onderzoekers te geleiden en dubbele invoer te vermijden en in tweede instantie om eerste stappen te zetten om deze metadata ook door te zetten naar de rest van het onderzoeksproject (o.a. door deze op te nemen in de CRIS systemen).

WAT WE WILLEN BEREIKEN

HOOFDDOELSTELLING

Inzetten op ketenintegratie door de voor Open Science relevante gegevens die in dit proces worden vastgelegd over onderzoeksprojecten te harmoniseren en in te zetten bij de uitvoering en afronding ervan. Dit laatste door het vermijden van dubbele invoeren van dezelfde gegevens en door 'het meenemen' van relevante metadata bij de uitvoering en afronding van het onderzoeksproject en door waar mogelijk automatisering toe te passen.

ENKELE CONCRETE VOORSTELLEN

Wat betreft de informatie met betrekking tot het 'meenemen' van Open Science informatie in de rest van de keten hieronder enkele mogelijkheden ter registratie:

- Informatie uit het DMP kan worden gebruikt om de onderzoeker te adviseren over geschikte repositories en licenties voor publicaties, datasets en software.
- Informatie over open/open onder condities/gesloten kan leiden tot automatische instelling van embargo's in het repository of een automatische validatie wanneer de gekozen voorwaarden afwijkend zijn van de gegevens uit de researchadministratie.
- De uit het researchadministratieproces voortvloeiende metadata kunnen worden doorgegeven en waar nodig aangevuld bij het archiveren en publiceren van onderzoek output zoals datasets, software of publicaties.

Wat betreft de ontwikkelde/in ontwikkeling zijnde researchadministratieplatformen:

- De huidige situatie is zeer gefragmenteerd, met tegelijkertijd veel instellingen die nog geen dergelijk platform ontwikkelen. Dit lijkt een kans om e.e.a. te coördineren, gebruik te maken van de *lessons learned* en de overwegen hoe e.e.a. het beste landelijk vormgegeven kan worden door op korte termijn een afsprakenstelsel over de Open Science aspecten te ontwikkelen en de krachten te bundelen om deze in te zetten voor de rest van het onderzoeksproces. Op de langere termijn zou dit kunnen leiden tot de ontwikkeling van een gezamenlijk research administratiesysteem.
- Brede invoering en koppeling van de persistent identifiers RAiD, Grant ID en ROR is een essentiële randvoorwaarde bij dit alles.

GAP ANALYSE

Volgens onze informatie zijn er verspreid over diverse instellingen enkele eerste stappen gezet bij de ontwikkeling van een researchadministratieplatform om ook metadata in te zetten voor de rest van het onderzoeksproces.

STAKEHOLDERS

Research offices bij universiteiten en faculteiten; ontwikkelaars van researchadministratiesystemen, NWO, ZonMW en andere onderzoeksfinanciers, bij onderzoek betrokken medewerkers zoals datastewards.

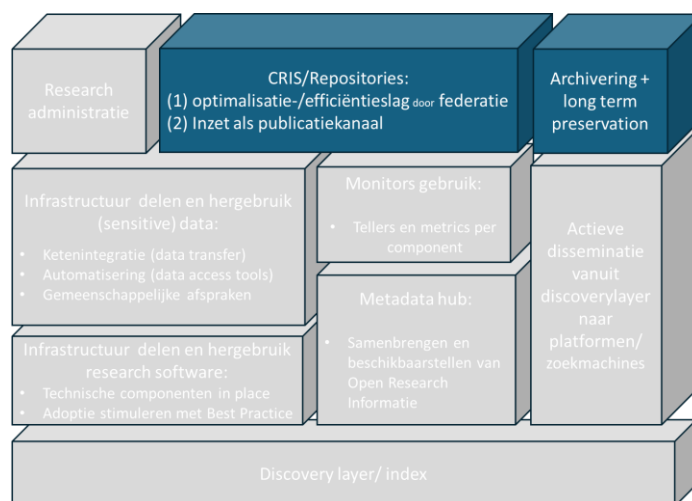
AANBEVELING

Gemeenschappelijke afspraken over harmonisatie van de Open Science gegevens ontwikkelen en de technische mogelijkheden in kaart brengen om vanuit deze systemen de rest van het onderzoeksproces wat betreft Open Science aspecten te voeden, daarbij gebruik makend van de *lessons learned* van de verschillende ontwikkelde/in ontwikkeling zijnde research administratiesystemen van de verschillende instellingen.

3.3.2 CRIS/REPOSITORIES: OPTIMALISATIE- EN EFFICIËNTIESLAG EN INZET ALS PUBLICATIEKANAAL

HUDIGE SITUATIE

- **Financiële zorgen en digitale soevereiniteit:** Er zijn zorgen over de financiële houdbaarheid van de oplopende kosten van Gold en Hybrid Open Access voor tijdschriftartikelen met Nederlandse co-auteurs (ca. 53.000 per jaar, waarvan ca. 60% met een Nederlandse *corresponding* auteur). Ook de afhankelijkheid van enkele grote uitgevers hierin speelt een rol met het oog op digitale soevereiniteit.
- **Innovatieve publicatievormen:** Mede ingegeven door het bovenstaande aandachtspunt is er een zoektocht gaande naar innovatieve Open Access publicatievormen:
 - Veel aandacht gaat uit naar Diamond OA tijdschriften met momenteel een beperkt volume van ca. 4% van de gepubliceerde artikelen.
 - Het publiceren van preprints krijgt eveneens meer aandacht (geschat op ca. 5000 per jaar), mede in het licht de loskoppeling van publiceren en peer review in het zogenaamde PRC-model. Een variant daarop wordt gevormd door o.a. Open Research Europe met *author-centered peer review*, waarbij de auteur na publicatie van het manuscript zelf twee peer reviewers uitnodigt.
 - Voor auteurs van niet-traditionele publicaties in het kader van Open Science – met name bij de start en bij de uitvoering van het onderzoek – zoals protocollen, modellen e.d. is er momenteel geen logisch publicatiekanaal.
- **Institutionele systemen en werkwijzen:** Veel instellingen hebben een institutioneel repository en een CRIS systeem ingericht. Deze zijn ingericht ten behoeve van registratie en rapportage (zoals de KUOZ rapportage) en betekenen een forse inspanning in termen van mensen en middelen. In de praktijk zijn deze systemen niet of nauwelijks op elkaar afgestemd en vindt er veel dubbel werk plaats. De metadata uit deze systemen wordt overigens wel geaggregeerd in PubliNova (voor hogescholen) en op landelijk niveau (NL Research Portal).



WAT WE WILLEN BEREIKEN

HOOFDDOELSTELLINGEN

- A. Optimalisatie en efficiëntieverbetering van de huidige CRIS/repository systemen
- B. Inzetten van deze op registratie en rapportage ingerichte systemen en inspanningen in het publicatieproces ten behoeve van (innovatieve) publicatievormen.

Relevant objective NPOS2030 Ambition Document and Rolling Agenda

3.4 In 2030, there will be a public platform for scientific publications, that allows everyone in the Netherlands to re-use global scholarly output for their professional and societal engagement practices.

CONCRETE VOORSTELLEN

De twee hierboven genoemde hoofddoelstellingen hebben we uitgewerkt in enkele concrete voorstellen die in beperkte kring besproken zijn. Hieruit is nog een aanvullend voorstel gekomen. Bij deze voorstellen dient aangetekend te worden dat deze meer discussie met een bredere kring van betrokken stakeholders behoeven.

1. Optimalisatie en efficiëntieverbetering van de huidige CRIS/repository systemen: Dit voorstel houdt een federatie in voor het invoeren van publicatie-objecten en de metadatering ervan in de CRIS/repositories (efficiëntie en ontduubeling) en een vorm van de JISC publication router om het harvesten van de full text zoveel mogelijk te automatiseren. Dit betekent o.m. dat output met meerdere auteurs van verschillende kennisinstellingen uitgewisseld wordt tussen de repositories/CRIS systemen op een zodanige manier dat er één DOI wordt toegekend, alle auteurs ORCIDs krijgen en dat er maar één instelling metadateert.

2A. Inzetten van de CRIS/repositories infrastructuur als publicatiekanaal: In het proces van outputregistratie → institutional repository → aggregatie, ligt de aandacht van de CRIS-systemen en institutionele repositories op registratie (achteraf) en het verzorgen van open toegankelijkheid. Met deze infrastructuuronderdelen verzorgen de instellingen al een deel van het publicatieproces waarmee wetenschappelijke output vindbaar wordt gemaakt. Met name voor output (zoals rapporten) die niet naar uitgevers gaat is dit proces al feitelijk de stap van intern document naar publieke beschikbaarheid zoals hierboven werd geschetst voor onderzoeksrapporten. In dit proces krijgen sommige documenttypen ook al een “institutional” branding. Deze “publicatiefunctie” kan veel krachtiger kan worden ingezet dan nu het geval is op de volgende wijze:

- Het voorstel is om een aggregator voor publicatieobjecten afkomstig uit de CRIS/repositories van de verschillende instellingen in te richten en dit in te zetten als publicatiekanaal ten behoeve van auteurs: door de serviceorganisaties rond CRIS/repositories (meestal bibliotheken) verrijkte metadata en de publicatie kunnen doorgezonden worden door middel van een ‘publicatiekanaal’ naar de betreffende preprintserver, Diamond of ander tijdschrift via het Notify mechanisme van COAR voor een overlay journal. Maar het publicatieobject kan ook gewoon blijven staan op de aggregator en daar eventueel door lezers via een Postpub review optie worden becommentarieerd. In principe zou dit ook kunnen vanuit de repositories (dus zonder aggregator), maar dan dienen er tientallen connecties gemaakt te worden en is het volume per repository te laag om interessant te zijn voor veel ‘traffic’ aan te trekken. Het *composability* argument en het volume argument pleiten voor een landelijk platform. Wanneer dit platform een succes is, kan het wel worden uitgebouwd met bijvoorbeeld ingangen op SDG’s, thema’s e.d.
- Overige functies van de publicatie aggregator: (1) Vanuit de aggregator kan met één connector de harvesting voor het eDepot van de KB plaatsvinden. (2) De aggregator kan een vertrekpunt

vormen om betekenisvolle toegang tot wetenschappelijke informatie te geven voor het brede publiek: een of meer publicatieobjecten kunnen worden samengevat in het Nederlands en/of lektentaal door AI-tools.

- De feedback op dit voorstel richtte zich voornamelijk op de haalbaarheid en de toegevoegde waarde van een dergelijke inrichting als publicatiekanaal voor de onderzoekers/auteurs, die dit ook als een inbreuk op de academische vrijheid zouden kunnen ervaren. Daarnaast zag men als moeilijkheid het versiebeheer van een publicatie in zijn verschillende gedaanten en vond men de relatie tussen het institutionele repository en het nationale platform onduidelijk.

2B. Institutionele CRIS systemen en een nationaal repository: Naar aanleiding van de discussie over het bovenstaande voorstel gaven meerdere respondenten uit de consultatiefase aan niet te hechten aan een institutionele repository en zich goed te kunnen voorstellen dat er een nationaal repository wordt ingericht dat de institutionele repositories vervangt. De metadatering wordt dan verzorgd vanuit de CRIS systemen, die wel institutioneel blijven. Een dergelijk nationaal repository zou weer enkele van de hierboven genoemde functies kunnen vervullen. Tenslotte werden hier mogelijkheden genoemd om zo'n platform ook geschikt te maken voor Diamond OA journals en werd er gewezen op recente ontwikkelingen rond Open Research Europe, die aangeven dat dit platform in de nabije toekomst open source wordt en naar verwachting ook additionele functionaliteiten zal krijgen.

Archivering en long-term preservation: De hier voorgestelde aggregator voor publicaties is tevens relevant voor de archivering en long-term-preservation: de publicaties kunnen d.m.v. één connector met het e-Depot van de KB aldaar worden gearchiveerd en gepreserveerd voor de langere termijn. Daarbij is overigens de URN:NBN-resolver van belang.

Beschikbaarstelling voor breed publiek: Ook kan de voorgestelde aggregator een startpunt zijn voor de beschikbaarstelling van onderzoekspublicaties op een voor breed publiek betekenisvolle wijze, bijvoorbeeld door middel van met AI vertalingen in het Nederlands en/of in niet-wetenschappelijke taal.

GAP ANALYSE

- **Gap 1: Zeer verschillende werkwijze en opnamebeleid wat betreft de CRIS/repositories van de instellingen:** Enkele illustraties van deze verschillen:
 - In één instelling worden de onderzoekers gevraagd om zelf hun publicaties in te voeren (onderzoekers ervaren dit als een bureaucratische last). Een andere instelling daarentegen ontzorgt de onderzoekers volledig door metadata over hun publicaties op te halen uit verschillende internationale bronnen (zoals Scopus en Open Alex).
 - Veel instellingen hebben overigens wel een vastgelegde, specifieke workflow voor dissertaties vanwege het promotiereglement.
 - Deze situatie leidt ertoe dat er verschillende records worden geproduceerd over dezelfde publicatie. Dit gebeurt *tussen* instellingen (een publicatie met auteurs van verschillende universiteiten of hogescholen) én zelfs soms *binnen* een instelling wanneer verschillende auteurs van dezelfde instelling hun publicaties invoeren in het CRIS systeem.
- **Gap 2: Er is een gap wat betreft niet-traditionele publicatietypen:**
 - Meerdere instellingen gaven aan geen actief opnamebeleid voor niet-traditionele publicaties te hebben ('Wij doen niks met grijze literatuur').

- Erkennen en waarderen speelt hierin een belangrijke rol. Nu er geleidelijk aan ook erkenning en waardering komt voor niet-traditionele publicatietypen zouden de CRIS/repository systemen daarop dienen in te spelen. In dat kader wordt opgemerkt dat er bij de universitaire medische centra een verkenning gaande is naar de mogelijkheden om preprints (die aan bepaalde kwaliteitscriteria voldoen) te erkennen als een voor een proefschrift geldige publicatie.

STAKEHOLDERS

Stakeholders zijn o.m. managers van CRIS/repository systemen, auteurs van preprints en andere niet-traditionele publicaties.

AANBEVELINGEN

Het voorstel is om de hieronder genoemde twee stappen in SPII Deel 2 verder uit te werken:

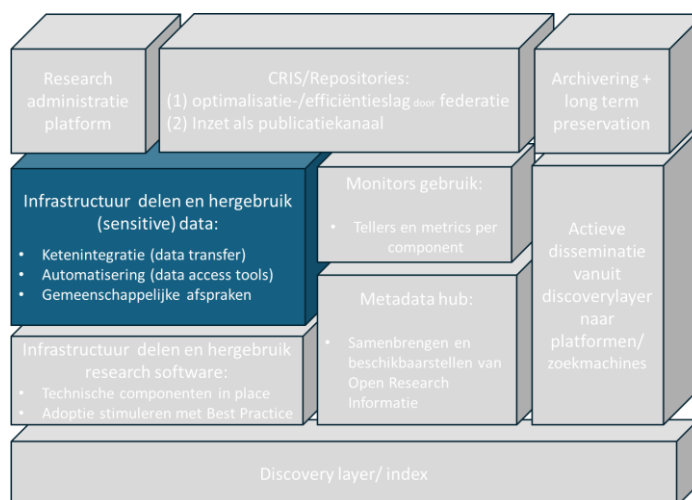
- CRIS/repository systemen:
 - Synchronisatie en harmonisatie van de werkwijzen en opnamebeleid
 - Federatie van deze systemen met het oog op efficiëntie en optimalisatie.
- Andere inrichting van de infrastructuur ten behoeve van het publiceren:
 - Verdere uitwerking van de mogelijkheden van een nationaal repository/publicatie aggregator/publicatieplatform voor met name de niet-traditionele publicaties en de innovatieve publicatievormen zoals het PRC model.

3.3.3 INFRASTRUCTUUR DELEN EN HERGEBRUIK (SENSITIEVE) DATA

HUDIGE SITUATIE

We onderscheiden de volgende componenten:

- **Datamanagement omgevingen en werkomgevingen:** Het gaat om *workspaces* waarin onderzoekers van een onderzoeksteam hun data kunnen verzamelen, bewerken en analyseren. Deze omgevingen ondersteunen ook de workflow van de onderzoekers en zijn gekoppeld aan opslagomgevingen waar de data opgeslagen worden. SURF Research Cloud^c en Yoda gelden hierbij als voorbeelden.
- **Veilige omgeving voor data analyse voor zeer gevoelige data:** een veilige omgeving met *compute to data* functionaliteit voor zeer gevoelige data (bijvoorbeeld SANE).
- **Datarepositories:** Datarepositories voor archivering en ontsluiting van de data (soms gekoppeld aan de hierboven genoemde workspaces). Hierbij zijn afspraken over datagovernance belangrijk (hoe lang worden data bewaard, wie is verantwoordelijk voor het toegangsbeheer e.d.). Voorbeelden zijn Dataverse.NL, en Yoda.
- **Long-term datarepositories:** Datarepositories voor archivering, ontsluiting én long-term preservation van data. Voorbeelden zijn 4TU.ResearchData en de DANS Data Stations.



Voor uitwisseling van datasets en hergebruik ervan spelen de volgende infrastructurele componenten^d een rol (al worden deze momenteel in de praktijk nog weinig gebruikt):

- **Data transfer tools**, waarmee datasets (veilig) overgedragen kunnen worden (tussen onderzoekers, van werkomgeving naar datarepository en van data repository naar werkomgeving)
- **Data-access mechanismen**, die de voorwaarden van de toegang tot een dataset regelen. Voorbeeld is Data Access Broker (DAB) dat researchers toegang geeft tot data providers d.m.v. een DOI (bij open access direct toegang, bij access onder voorwaarden, verwijzing naar de landing page van de data provider).

^c [SURF Research Cloud](#) is een portal waarin virtuele onderzoeksomgevingen gebouwd kunnen worden met o.m. voorgeconfigureerde workspaces en datasets gebruiken of deze zelf toevoegen. Partijen kunnen bijdragen aan de functionaliteit door hun reken- en datadiensten te integreren.

^d Daarnaast zijn er encryptie en pseudonimisatie tools.

- **Toegangsbeheermechanismen:** toegang tot data én data management omgevingen voor individuele onderzoekers waarmee wordt samengewerkt (zie SRAM - SURF Research Access Management).

Het is van belang op te merken dat de ontwikkeling van de infrastructuur voor research data management parallel loopt met de Open Science beweging m.b.t. de FAIR criteria en de reproduceerbaarheids crisis, waardoor het onderscheid 'research infrastructuur' en 'Open Science infrastructuur' in discussies snel door elkaar loopt. De insteek van SPII ligt in het Open Science perspectief op de RDM infrastructuur. In dat licht constateren we het volgende:

- Een grote diversiteit aan RDM-infrastructurele componenten, met een landschap dat nog volop in ontwikkeling is en waarin het mede daardoor lastig is voor onderzoekers om een weg te vinden.
- Geen of onvoldoende integratie van de keten research administratie → data management plan → data management omgeving → (long-term) data repository
- Ontbreken van instellingsbrede afspraken t.a.v. data governance (wie is verantwoordelijk voor toegang, voor bewaartermijnen etc.)
- Lacunes in automatisering: een data-access mechanisme dat het toegang geven aan datasets onder condities automatiseert (voorbeeld: Research Data Exchange (RDX) dat onvoldoende uitontwikkeld is^e).

WAT WE WILLEN BEREIKEN

1. Ketenintegratie waar mogelijk door middel van federatie
2. Automatisering waar mogelijk om de onderzoekers te ontlasten
3. Gemeenschappelijke afspraken die (1) en (2) mogelijk maken.

Relevant objective NPOS2030 Ambition Document and Rolling Agenda

4.4 In 2027, Dutch FAIR research services and infrastructure are efficiently organised and connected, closely linked to the European digital infrastructure. There are FAIR data service infrastructures based upon major data sharing initiatives (e.g. SURF, DANS), including a network of FAIR data portals as part of EOSC.

GAP ANALYSE

De volgende gaps zijn geïdentificeerd:

- **Gap bij research administratie:**
 - In paragraaf 3.3.1 is al ingegaan op het proces van research administratie. Wat betreft datasets gaat het om die metadata in het rest van het onderzoeksproject 'mee te nemen'

^e RDX houdt het volgende in: (1) de onderzoeker stelt de dataset beschikbaar onder condities door middel van een aantal vragen; (2) de potentiële gebruiker tekent voor de toegang digitaal een contract, waarbij hij/zij zich verplicht om aan de condities te voldoen.

die relevant zijn voor het delen en hergebruik van onderzoeksdata en voor de reproduceerbaarheid van het onderzoek. Dit wordt gezien als een grote kans: een koppeling met research administratie die het mogelijk maakt om de tijdens dat proces ontwikkelde metadata mee te nemen, te hergebruiken en verder aan te vullen.

- **Gap m.b.t. DMP:**

- DMP's zijn momenteel vooral statische documenten en hebben het karakter van 'bezint eer ge begint' en soms ook het karakter van een contract. Het is echter wenselijk om het ook als een bron van informatie te gebruiken voor de rest van het onderzoeksproject. In dit verband wordt gesproken over de mogelijke functies van machine-readable DMP's hierbij^f.

- **Gap bij overgang van uitvoering naar archiveren/publiceren van een dataset:**

- Het is wenselijk dat de metadata van onderzoeksdatasets in uitvoering automatisch worden meegenomen wanneer de onderzoekers de datasets archiveren en/of publiceren via een (long term) data repository. Daarmee wordt de informatie tijdig vastgelegd en wordt redundante invoer voorkomen.

- **Gap bij (de automatisering van) data access mechanismen:**

- Voorbeelden van geautomatiseerde data access mechanismen zijn de Data Access Broker dat onderzoekers verbindt met providers [d.m.v. een DOI: bij open access direct toegang, bij access onder voorwaarden, de landing page van de data provider] en Research Data Exchange (RDX) dat beide kanten van het proces regelt (maar onvoldoende uitontwikkeld is). Hierbij stelt de onderzoeker de dataset beschikbaar onder condities door middel van een aantal vragen en tekent de potentiële gebruiker voor de toegang een contract met de gebruikscondities. Op zo'n manier worden data providers ondersteund om controle te houden over wie gevoelige data gebruikt en op welke manier, en wordt het aantrekkelijk voor onderzoekers om gevoelige data FAIR te delen.

- **Gap in landelijke afspraken over onder meer data governance:**

- Het gaat om landelijke afspraken over onder meer data governance, zoals het regelen van verantwoordelijkheden voor toegang (met name bij de overdracht aan een datarepository), maar ook om het harmoniseren van instellingsbeleid (bijvoorbeeld met betrekking tot de bepaling van sensitiviteit van research data).

OVERZICHT STAKEHOLDERS

De betrokken stakeholders zijn onder meer data stewards, beheerders en ontwikkelaars van data management omgevingen en data repositories, DCC's en TDCC's, SURF, 4TU.ResearchData en DANS.

AANBEVELING

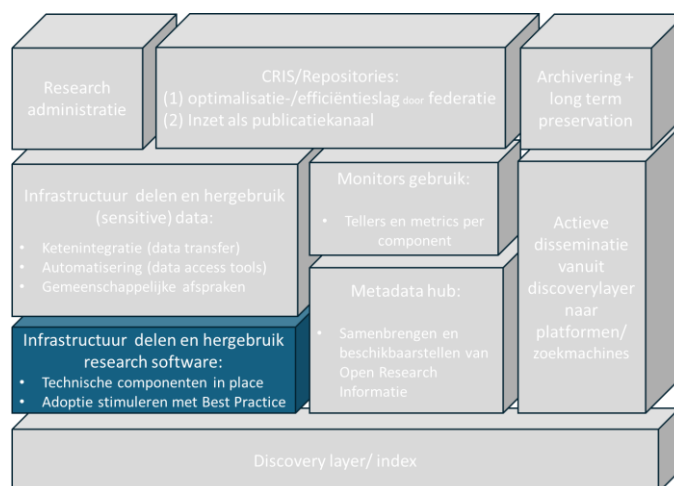
^f Er lijkt qua tools voor het opstellen van DMP's een zekere consolidatie zichtbaar rond enkele tools. Daarbij werden Argos, Data steward wizzard en DMP online genoemd. DCC-PO heeft voor de (36) hogescholen een generiek DMP-plan ontwikkeld. Dit zijn mogelijke aanknopingspunten voor het ontwikkelen van machine-readable DMP's met het oog op het 'meenemen' van metadata in de rest van het onderzoeksproces.

- Ontwikkel een samenhangende aanpak van Research Data Management over de volledige data life cycle, van projectplanning tot hergebruik, in samenwerking met de stakeholders. Wat betreft de Open Science aspecten gaat het daarbij om (1) de tijdens de verschillende stappen ontwikkelde metadata die relevant zijn voor het delen en hergebruik van de data en voor de reproduceerbaarheid van het onderzoek tijdens het onderzoeksproces mee te nemen, te hergebruiken en verder aan te vullen en (2) om de genoemde gaps bij de automatisering van data access mechanismen aan te pakken. SPII Deel 2 zou de Open Science aspecten kunnen aanpakken, in nauwe afstemming met SURF, dat een coördinerende rol heeft bij het verbinden en/of ontwikkelen van de researchinfrastructuur van RDM componenten.

3.3.4 INFRASTRUCTUUR DELEN EN HERGEBRUIK SOFTWARE

HUDIGE SITUATIE

Sinds enkele jaren is de Research Software Directory (RSD) ontwikkeld door het eScience Center. Doel is om hergebruik van research software te stimuleren, te verbinden met andere research output en te zorgen voor het citeren van softwaregebruik met het oog op erkenning en waardering van de betreffende onderzoekers en research software engineers. Naast implementatie van de RAiD identifier en de SPDX identifier en de metadatastandaarden CFF en Codemeta is met de inzet van de RSD alles in technische zin aanwezig voor wat betreft archivering van software (in Zenodo of Software Heritage) met het oog op linken aan publicaties.



WAT WE WILLEN BEREIKEN

Adoptie van de bovengenoemde infrastructurele componenten door middel van het ontwikkelen van Best Practices voor Research Software.

Relevant objective NPOS2030 Ambition Document and Rolling Agenda

4.7 In 2030, the Dutch open science community has access to a robust, cost-effective, and sustainable research software ecosystem, which allows for publication and archiving of software in online software repositories (e.g. GitHub, Gitlab) by RPOs, as well as integration and interoperability with existing national, international, and domain-oriented archives and Current Research Information Systems.

GAP ANALYSE

De gap is dat de technische componenten operationeel zijn, maar dat het schort aan de adoptie ervan.

OVERZICHT STAKEHOLDERS

Research software engineers, onderzoekers, eScience, TDCC's, DCC's.

AANBEVELING

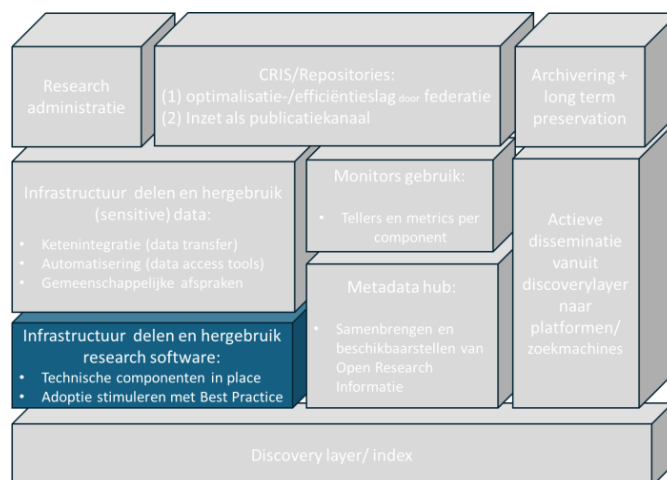
Aanbevolen wordt om Best Practices te ontwikkelen voor het FAIR archiveren en publiceren van research software. Kort na het einde van SPII Deel 1 zullen de resultaten worden gepubliceerd van een TDCC NES project met de titel '[Best Practices for Sustainable Software](#)'. Mogelijk dat hiermee deze aanbeveling al is ingevuld of dat er voor andere vakgebieden aangepaste Best Practices dienen te worden ontwikkeld.

3.3.5 OPEN RESEARCH INFORMATIE, DISCOVERY EN ACTIEVE DISSEMINATIE

HUDIGE SITUATIE

Een belangrijk deel van de research informatie over Nederlands onderzoek is niet open beschikbaar. Het deel dat wel open is, is gefragmenteerd beschikbaar. De [Barcelona Declaration](#) – door meerdere stakeholders in Nederland ondertekend – roept onder meer op tot het open maken van research informatie en hiervoor duurzame infrastructuur te ontwikkelen.

De discovery en disseminatie van de Nederlandse onderzoeksresultaten wordt sinds 2023 gerealiseerd door de aggregatie van de metadata vanuit de institutionele repositories in de [Netherlands Research portal](#) als onderdeel van OpenAIRE.



WAT WE WILLEN BEREIKEN

HOOFDDOELSTELLINGEN

- Een infrastructuur voor de beschikbaarstelling van open research informatie over Nederlands onderzoek met een zo compleet mogelijke dataverzameling die door alle partijen gebruikt en geanalyseerd kan worden.
- Verbetering van de discovery en disseminatie van de Nederlandse onderzoeksresultaten met als resultaat een verhoogde (inter-)nationale zichtbaarheid.

Relevant objective NPOS2030 Ambition Document and Rolling Agenda

3.2 In 2027, the digital infrastructure for scientific publications and metadata is sustainable and preferably open, and supports publication, (re-)use and analysis by anyone. This includes the development of a national publication platform and an Open Knowledge Base, as well as support for existing open archives, repositories and infrastructures (such as Publinova) for publication, data, software and other scholarly outputs.

CONCRETE VOORSTELLEN

- **Tellers en metrics:** Op de verschillende componenten van OSI dienen tellers worden geïnstalleerd voor het tellen van downloads en views op publicaties, en het meten van het hergebruik van datasets (bijvoorbeeld door [Make Data Count](#)) en research software. Aansluiting bij de in aanbesteding zijnde Open Science Monitor in Nederland en de internationale initiatieven hierin ([Open Science Monitoring Initiative](#)) dient hierbij te worden gezocht.
- **Metadatahub:** De metadatahub brengt de informatie over het onderzoek én de metrics bijeen door deze te verzamelen uit de verschillende componenten van het systeem, waar mogelijk te koppelen, en stelt deze als Open Research Information beschikbaar.

- **Discovery layer:** De discovery layer bevat dezelfde informatie als de metadatahub, maar presenteert deze op een andere wijze die geschikt is voor discovery.
- **Actieve disseminatie:** De informatie in de discovery layer wordt d.m.v. actieve disseminatie verspreid naar diverse (inter-)nationale zoekmachines, aggregatie-platformen e.d. Dit moet leiden tot een verhoogde zichtbaarheid van het Nederlandse onderzoek.

GAP ANALYSE

Een gedeelde infrastructuur met open research informatie over het Nederlandse onderzoek ontbreekt momenteel met als gevolg fragmentatie (verschillende actoren verzamelen en verwerken onderzoeksinformatie op verschillende manieren). Een dergelijke infrastructuur – hier metadatahub genoemd – dient de volgende gegevens te verzamelen:

(1) Research input:

- Het gaat om gegevens over o.m. personen, projectsubsidies, data management plannen, apparatuur. Deze informatie dient verzameld te worden uit institutionele systemen zoals researchadministratieplatformen en CRIS-systemen en uit systemen van partijen als NWO, CrossRef en d.m.v. identifiers als Grant ID en RAID).
- Hier wordt een belangrijke gap gesignaleerd omdat deze informatie in de praktijk lastig te verzamelen blijkt te zijn omdat bij veel instellingen deze informatie niet op orde is.

(2) Research output:

- Het gaat o.m. om gegevens over publicaties, datasets, software, patenten, evenementen (media optredens).
- Deze informatie is vaak wel te verzamelen maar verdeeld over vele bronnen.

(3) Research outcome:

- Het gaat om gegevens zoals views, downloads, forks, wetenschappelijke citaties in peer-reviewed artikelen, en vermeldingen in sociale media, nieuwsartikelen en beleidsdocumenten.
- Deze informatie is verspreid over vele bronnen en deels in handen van commerciële aanbieders.

Verrijking: Voorzien wordt dat de metadatahub mechanismen heeft voor kwaliteitsbewaking van de op te nemen gegevens en met tools de verzamelde gegevens verrijkt (bijvoorbeeld modules die m.b.v. tekst en datamining informatie uit de full text van een artikel halen over bijvoorbeeld de financier van het onderzoek, SDG's en dergelijke).

De Netherlands Research portal heeft (nog) lacunes wat betreft de kwaliteit en dekkingsgraad van de repositories (zie paragraaf 3.3.1) en wat betreft de functionaliteit van de portal (zie de [roadmap](#)).

AANBEVELING

- Zet een metadatahub op die informatie over het onderzoek én de metrics bijeenbrengt door deze te verzamelen uit de verschillende componenten van het systeem, waar mogelijk te koppelen, en deze als Open Research Information beschikbaar stelt.
- Gebruik deze informatie om d.m.v. een discovery laag de informatie over Nederlands onderzoek actief te verspreiden.

3.3.6 FEDERATIEF AFSPRAKENSTELSEL, AANVULLENDE FUNCTIONALITEIT, ONTWERPCRITERIA EN RANDVOORWAARDEN

FEDERATIEF AFSPRAKENSTELSEL

Een federatief afsprakenstelsel is essentieel om de federatieve opzet van OSI mogelijk te maken, te onderhouden en verder te ontwikkelen. Zoals eerder gezegd, gaat dit om afspraken op de volgende vlakken:

- Technisch en semantisch:
 - PIDs: dit betreft de consistente applicatie van persistent identifiers door alle aan de federatie deelnemende partijen. In de federatieve schema's zijn een aantal PIDS genoemd die min of meer standaard zijn (DOI, ORCID, NBN, SWID), maar vaak nog inconsistent worden ingezet. Daarnaast worden er een reeks PIDs noodzakelijk geacht, waarvan de implementatie nog in de beginfase is: RAiD (voor projecten), GrantID's (voor projectsubsidies), ROR (voor organisaties), SPDX identifier (voor software licenties). Dit pleit voor een uitbreiding van de scope en een versnelde implementatie van het SURF project NL-PID-implementatie).
 - Metadata: afspraken over het gebruik van metadata-standaarden vormen eveneens een onderdeel van het afsprakenstelsel. Het gaat om onder meer de metadata-standaarden Marc21, CFF, Codemeta en het gebruik van gemeenschappelijke vocabulaires/ontologiën voor de beschrijvingen van de metadata. Daarbij moet opgemerkt worden dat het niet per se noodzakelijk is voor een federatieve connectie dat de twee systemen dezelfde metadata standaard hanteren. Het is mogelijk om een vertaling (crosswalk) in de connectie (meestal een API) in te bouwen naar een ander format/standaard, zodat de beide systemen adequaat met elkaar kunnen communiceren. Bij doorontwikkeling van de systemen kiest men er dan vaak voor om voortaan de 'andere format/standaard' te gaan hanteren.
 - Daarnaast zijn afspraken nodig over het harmoniseren van het opnamebeleid voor CRIS/repository systemen, definities van items, gemeenschappelijke templates voor contextuele documentatie en toe te passen datamodellen.
- Juridisch:
 - Een deel van het federatieve afsprakenstelsel ligt meer op het juridische vlak. Deze betreffen o.m. afspraken over eigenaarschap van de data, over het harmoniseren van de criteria voor sensitieve datasets (nu vaak in vier categorieën), afspraken over encryptie en decryptie procedures, data transfer/sharing agreements en over verantwoordelijkheden t.a.v. toegang en data governance.
- Organisatorisch:
 - Organisatorische kwesties betreffende de federatie (o.a. governance, operationele organisatie, support, opleidingen) zullen worden behandeld in SPII Deel 3.

AANVULLENDE FUNCTIONALITEITEN, ONTWERPCRITERIA EN RANDVOORWAARDEN

Naast de hierboven voorgestelde componenten en functies van OSI zijn er enkele aanvullende functionaliteiten, ontwerpcriteria en randvoorwaarden die essentieel zijn voor een succesvolle implementatie:

- **Gebruikersvriendelijkheid en workflow ondersteuning:** Onderzoekers en onderzoeksondersteunende professionals dienen optimaal ondersteund te worden in een workflow. Dit betekent dat interfaces intuïtief zijn, processen zoveel mogelijk gestroomlijnd en waar mogelijk handelingen geautomatiseerd plaatsvinden.
- **Dataopslag:** De infrastructuur die te voorzien in veilige en flexibele opslagmogelijkheden voor data, zowel open datasets als voor gesloten of beperkt toegankelijke datasets.
- **Kwaliteitscontrole:** Bij de publicaties in de institutionele repositories wordt een administratieve controle uitgevoerd door de serviceorganisaties (o.a. metadata, licenties e.d.). Een aanvullende mogelijkheid voor instellingen is om ook een academische controle uit te voeren om de kwaliteit van de geüploade content te waarborgen.
- **Metrics en reviewopties:** De eerder genoemde metrics en de bij de nationale publicatieaggregator genoemde Postpub review optie (commentaren van lezers) zijn ook van belang in het kader van het erkennen en waarderen van niet-traditionele onderzoekspublicaties, zoals protocollen, rapporten e.d. die daardoor aan status winnen.

Tot slot dient OSI dient te voldoen aan de randvoorwaarden zoals weergeven in paragraaf 2.3.3.

3.4 NAAR EEN INTEGRALE REFERENTIEARCHITECTUUR VOOR OSI

In dit project zijn vertrekpunten geformuleerd voor de ontwikkeling van een integrale referentiearchitectuur voor OSI. Deze vertrekpunten dienen in een wijder verband te worden bediscussieerd en verder uitgewerkt, vandaar dat ze hier in Bijlage C zijn weergegeven en niet in de hoofdttekst. Het is echter van belang te benadrukken dat van een dergelijke verder te ontwikkelen referentiearchitectuur voor OSI een belangrijke convergerende werking zal uitgaan. In Bijlage C staat ter illustratie de convergerende werking van de referentiestructuur voor digitaal erfgoed (DERA) in een landschap van erfgoedinstellingen geschetst dat qua diversiteit en decentrale opzet overeenkomsten vertoont met het landschap waarin OSI zal worden ingebed. Het voorstel is om de OSI referentiearchitectuur te zien als een eerste stap naar een bredere referentiearchitectuur voor onderzoek (HOSA voor onderzoek).

4. DISCUSSIE, CONCLUSIES EN AANBEVELINGEN VOOR SPII DEEL 2

4.1 DISCUSSIE EN EVALUATIE SPII DEEL 1

PROCES

PROJECTGROEP ALS DENKTANK

Het voor u liggende Strategisch Plan Integrale Infrastructuur voor Open Science Deel 1 met als ondertitel 'Contouren van een Open Science Infrastructuur' is ontwikkeld door de Projectgroep SPII Deel 1. Deze projectgroep bestond uit een kleine groep experts, die met veel enthousiasme hieraan gewerkt hebben en daarvoor vrijgemaakt waren door hun instellingen. De projectgroep heeft gefungeerd als een denktank en in korte tijd (24 maart tot 16 juli 2025) en met beperkte middelen een voorstel voor een Open Science Infrastructuur ontwikkeld door middel van een methode (use-cases → federatieve schema's → voorstellen voor federatieve infrastructurele componenten). Deze methode is beschreven in paragraaf 3.2 en Bijlage A.

BEPERKTE CONSULTATIE ROND VERSIE 0.7

Begin juni zijn de eerste resultaten beschreven in versie 0.7, die vervolgens is besproken in een consultatieronde met een - vanwege de tijdsdruk beperkt - aantal stakeholders. De ontvangen feedback is óf verwerkt in de voorliggende versie óf weergegeven in de discussie hieronder met suggesties om dit in SPII Deel 2 op te pakken.

SPII DEEL 1 ALS EERSTE STAP

Gegeven de beperkte tijd en middelen van de projectgroep zijn wij er ons er van bewust dat SPII Deel 1 meer detaillering, uitwerking en uitbreiding behoeft en naar verwachting op een aantal onderdelen na een meer uitgebreide consultatie aangepast zal worden. Dit rapport dient dan ook beschouwd te moeten worden als een eerste stap op weg naar een volwaardig strategisch plan. Ondanks de genoemde beperkingen hopen we dat dit plan voldoende aanknopingspunten bevat voor een vervolg. Zie onze voorstellen voor een SPII Deel 2 in paragraaf 4.3.

FEEDBACK UIT DE CONSULTATIERONDE

RELATIE OSNL PROJECTEN

In de consultatieronde kwamen er veel vragen over de relatie tussen dit plan en de bij OSNL ingediende projecten voor de call Open Science infrastructuur. Het antwoord ligt in de door de projectgroep gevolgde, onafhankelijke methode. Het feit dat sommige voorstellen voor de Open Science infrastructuur sterke gelijkenis tonen met voor de OSNL call ingediende projecten maakt naar onze mening zowel de gevolgde methode als deze projectvoorstellen krachtiger.

BEPERKTE CONSULTATIE

Een ander kritiekpunt was de beperkte consultatie, waarbij velen aangaven intensiever bij de ontwikkeling van SPII betrokken hadden willen worden.

Dit zien we als een zeer positief teken: er is duidelijk veel belangstelling voor een dergelijk strategisch plan en intensieve deelname van veel stakeholders aan de verdere ontwikkeling ervan in het kader van een vervolgaanpak lijkt dan ook gegarandeerd.

USE-CASES REFLECTEREN ONVOLDENDE DE GEHELE ONDERZOEKSCYCLUS

Een ander genoemd aandachtspunt zijn de use-cases waarop de federatieve schema's zijn gebaseerd. Gesteld werd dat er andere use-cases niet uitgewerkt zijn, zoals use-cases die relevant zijn voor onderzoekers tijdens de uitvoering van hun onderzoek. Daardoor ligt de nadruk van het rapport nu volgens deze zegslieden te veel op het delen van onderzoeksresultaten aan het eind van het project en is er te weinig aandacht voor de Open Science aspecten bij de uitvoering van het onderzoek.

In Figuur 8 (in Bijlage A) staan de use-cases gekoppeld aan de onderzoekscyclus. Dit laat zien dat er inderdaad geen use-cases zijn uitgewerkt voor de uitvoeringsfase van een onderzoekproject. De verwachting is dat dergelijke use-cases een sterk disciplinair karakter zullen hebben. In SPII Deel 2 dient bekeken te worden of er use-cases voor de uitvoeringsfase door stakeholders aangedragen worden en in hoeverre hieruit additionele gewenste componenten voor de federatieve infrastructuur uit voortvloeien. Wij zien dit als een bekrachtiging van de gevolgde methode. Uit een koppeling van de use-cases aan de onderzoekscyclus blijkt inderdaad dat er weinig use-cases zijn uitgewerkt voor de uitvoeringsfase van onderzoek. We onderschrijven dus deze beperking van de use-cases en pleiten ervoor om in de vervolgaanpak additionele use-cases – aan te dragen door stakeholders – op te nemen en daarop dezelfde methodiek (use cases → federatief schema → infrastructureel voorstel) toe te passen.

INFRASTRUCTURELE VOORSTELLEN EN VREES VOOR ZELFBOUW

Andere deelnemers aan de consultatieronde vonden de voors en tevens van de concretere voorstellen voor infrastructuur in dit rapport onvoldoende uitgewerkt en doorgesproken met de betreffende stakeholders. Als voorbeeld kan hier de voorgestelde publicatieaggregator op nationaal niveau gelden. Er is een vrees – mede gebaseerd op eerdere ervaringen – dat dit kan leiden tot bouw van een infrastructuur die in de praktijk niet of weinig gebruikt zal worden.

Zelfbouw van de publicatieaggregator lijkt ons onnodig met het oog op onder andere recente ontwikkelingen rond het open-source maken van het platform van Open Research Europe.

LOSKOPPELING VAN BESTAANDE SYSTEMEN

Enkele deelnemers bepleiten een meer functionele benadering van OSI componenten met een beschrijving in termen van data platformen, data producten, data management en tooling (o.a. m.b.v. AI). Anders gezegd: wegblijven van benamingen van bestaande systemen die immers aan veranderingen onderhevig zijn. Een dergelijke benadering zou duidelijker maken dat bepaalde modules ingezet kunnen worden voor meerdere functies binnen OSI. Wij onderschrijven dit en zien de ontwikkeling van een OSI referentiearchitectuur als manier om dit te bereiken.

4.2 CONCLUSIE

Uit de reacties op versie 0.7 van dit plan blijkt dat er veel belangstelling is onder de stakeholders voor een visie en strategie voor het ontwikkelen van een integrale, samenhangende infrastructuur voor Open Science. We hopen dat dit rapport gezien wordt als een eerste stap daartoe én als aanleiding om e.e.a. verder te ontwikkelen. Daarbij willen we benadrukken dat naar ons oordeel de kracht van het rapport ligt in de gevolgde methode en pleiten ervoor om deze methode – eventueel uitgebreid met meer use-cases – ook in te zetten voor SPII Deel 2. Voor SPII Deel 2 bevelen we in de volgende paragraaf een aanpak aan.

4.3 AANBEVELINGEN VOOR SPII DEEL 2

SPLITSING IN DEEL 2 EN 3

Het voorstel is om SPII Deel 2 op te knippen in twee delen 2 en 3:

- 2 met een technische insteek met als gewenst resultaat concrete voorstellen voor OSI
- 3 met een organisationele insteek rondom de concrete OSI voorstellen.

Bij de aanvang van SPII Deel 2 stellen we voor naar de suggesties in Bijlage D mee te nemen in de precieze opzet en scope van SPII Deel 2.

VOORSTEL VOOR WERKPROGRAMMA SPII DEEL 2



Figuur 7 Overzicht voorstel werkprogramma SPII Deel 2

Werkpakket A gericht op de technische uitwerking van de koppelvlakken van OSI:

- A. Ontwikkeling referentiearchitectuur
- B. Ontwikkeling metadata schema
- C. Eventuele uitbreiding en versnelling PID implementatie project (uitgevoerd door SURF Innovatiezone Open Science Versterken)
- D. Ontwikkeling API's en koppelingen discipline-gerichte research infrastructuur
- E. Koppeling met EOSC node

E.e.a. te ontwikkelen in nauwe samenhang met de hieronder weergegeven werkpakketten B t/m E.

Werkpakket B Technische uitwerking gericht op de voorgestelde OSI federatieve oplossingen in samenwerking met de betreffende stakeholder communities met de volgende deelwerkpakketen

A. Research administratie: Uitwerking van de aanbeveling:

Gemeenschappelijke afspraken over harmonisatie van de Open Science gegevens ontwikkelen en de technische mogelijkheden in kaart brengen om vanuit deze systemen de rest van het onderzoeksproces wat betreft Open Science aspecten te voeden, daarbij gebruik makend van de *lessons learned* van de verschillende ontwikkelen/in ontwikkeling zijnde research administratiesystemen van de verschillende instellingen.

B. CRIS/repositories: optimalisatie- en efficiëntieslag en inzet als publicatiekanaal: Uitwerking van de aanbevelingen:

- CRIS/repository systemen:
 - Synchronisatie en harmonisatie van de werkwijzen en opnamebeleid
 - Federatie van deze systemen met het oog op efficiëntie en optimalisatie.
- Andere inrichting van de infrastructuur ten behoeve van het publiceren:
 - Verdere uitwerking van de mogelijkheden van een nationaal repository/publicatie aggregator/publicatieplatform voor met name de niet-traditionele publicaties en de innovatieve publicatievormen zoals het PRC model.
 - Connector publicatieaggregator en KB eDepot
 - Publicatieaggregator als startpunt voor beschikbaarstelling voor breed publiek

C. Infrastructuur delen en hergebruik (sensitieve) data: uitwerking van de aanbevelingen:

- Ontwikkel een samenhangende aanpak van Research Data Management over de volledige data life cycle, van projectplanning tot hergebruik, in samenwerking met de stakeholders. Wat betreft de Open Science aspecten gaat het daarbij om (1) de tijdens de verschillende stappen ontwikkelde metadata die relevant zijn voor het delen en hergebruik van de data en voor de reproduceerbaarheid van het onderzoek tijdens het onderzoeksproces mee te nemen, te hergebruiken en verder aan te vullen en (2) om de genoemde gaps bij de automatisering van data access mechanismen aan te pakken.
- Aanpak van de Open Science aspecten door SPII Deel 2 in nauwe afstemming met SURF, dat een coördinerende rol heeft bij het verbinden en/of ontwikkelen van de researchinfrastructuur van RDM componenten.

D. Infrastructuur delen en hergebruik research software: uitwerking van de aanbeveling:

- Ontwikkel Best Practices voor het FAIR archiveren en publiceren van research software in aanvulling op de resultaten van het TDCC NES project '[Best Practices for Sustainable Software](#)'.

E. Metadata hub voor ORI en Discovery layer en actieve disseminatie: uitwerking van de aanbevelingen:

- Zet een metadatahub op die informatie over het onderzoek én de metrics bijeen door deze te verzamelen uit de verschillende componenten van het systeem, waar mogelijk te koppelen, en stelt deze als Open Research Information beschikbaar.
- Gebruik deze informatie om d.m.v. een discovery laag de informatie over Nederlands onderzoek actief te verspreiden.

Werkpakket C: Prioritering, ontwikkeling roadmap, ontwikkeling businesscases en kostenbatenanalyses voorgestelde ‘data-producten’

We voorzien een derde werkpakket dat – op basis van de resultaten van het werkpakketten A en B een prioritering vaststelt voor de op te stellen roadmap, 2 of 3 dataproducten, zoals uitgewerkt in werkpakket B, uitgewerkt wat betreft kosten-batenanalyse en businesscase en op basis daarvan deliverable 3 (zie hieronder) oplevert.

Deliverables:

(1) Technische infrastructurele elementen:

- Referentiearchitectuur/streefarchitectuur
- Metadataschema
- Technische standaarden en afspraken over koppelvlakken, inclusief koppeling EOSC-NL node

(2) Voorstel voor een federatief afsprakenstelsel

- Voorstel technische en semantische afsprakenstelsel voortkomend uit de bovengenoemde werkpakketten
- Voorstel juridische afspraken voortkomend uit de bovengenoemde werkpakketten.
- Relatie met afspraken t.a.v. EOSC-NL node afgestemd.

(3) Roadmap OSI en voorstellen voor ontwikkeling en implementatie 2026-2029

- Voorstellen voor in de periode 2026 – 2029 te ontwikkelen van enkele ‘data producten’ die een hoge prioriteit hebben voor OSI, inclusief businesscase en kosten-batenanalyse.
- Schets roadmap OSI voor de langere termijn

Uitvoering:

- Door een te formeren projectgroep waarbij projectgroepleden de coördinatie op zich nemen van een werkpakket en stakeholders vragen om deel te nemen aan werkgroepen en/of workshops rond een werkpakket.
- Indien mogelijk is het voorstel om alle communicatie van het project openbaar te maken, waarbij iedereen die dat wil kan deelnemen of commentaar leveren, volgens het voorbeeld van [W3C](#) (dit vereist wel een omgeving waarin dat mogelijk is).

SPII DEEL 3

Wanneer de deliverables van SPII Deel 2 omarmd worden, kunnen de organisationele aspecten rond deze deliverables ontwikkeld worden. E.e.a. in nauwe afstemming met de organisationele ontwikkelingen rond de EOSC-NL node. Daarbij dient gedacht te worden aan het volgende:

- Governance structuur
- Operationele organisatie
- Financiering.

5. REFERENTIES

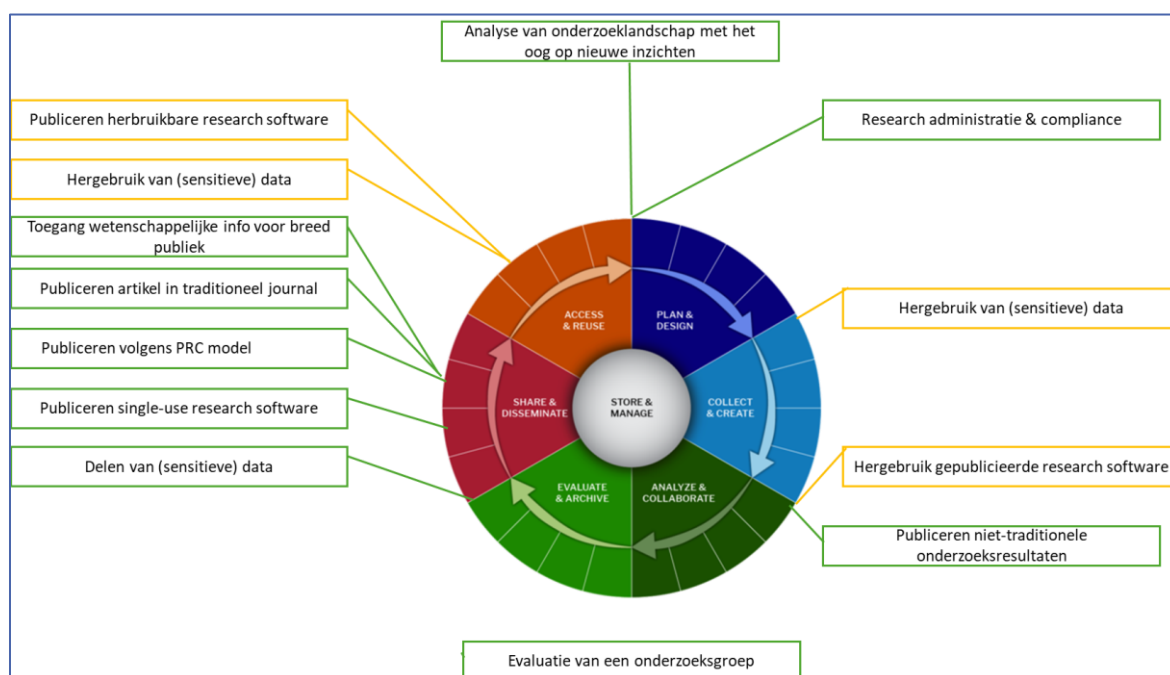
1. Oort, F. & Schreurs, E. Integrated Infrastructure for Open Science. (2025); <https://zenodo.org/records/10954657>
2. Bilder, G., Lin, J. & Neylon, C. Principles for Open Scholarly Infrastructures-v1. (2015) doi:10.6084/m9.figshare.1314859.v1.
3. NPOS. Open Science 2030 in the Netherlands: NPOS2030 Ambition Document and Rolling Agenda. <https://zenodo.org/record/7433767> (2022) doi:10.5281/ZENODO.7433767.
4. CARE Principles. Global Indigenous Data Alliance <https://www.gida-global.org/care> (2023).
5. Open Science agenda UNL; <https://www.universiteitenvannederland.nl/files/publications/Open%20Science%20agenda%20UNL.pdf>.
6. Innovatiezone Open Science Versterken. (2024). <https://www.surf.nl/innovatiezone-open-science-versterken>
7. Open Science Infrastructure | Open Science NL. <https://www.openscience.nl/en/calls/open-science-infrastructure> (2024).
8. Werkagenda: Samen waardengericht onderweg | SURF.nl. <https://www.surf.nl/themas/publieke-waarden/werkagenda-samen-waardengericht-onderweg>.
9. Brembs, B. et al. Plan I - Towards a sustainable research information infrastructure. (2021) doi:10.5281/zenodo.4454640.
10. Goudarzi, S. & Dunks, R. Defining Open Scholarly Infrastructure: A Review of Relevant Literature.
11. Woodford, C. et al. The Global Open Research Commons International Model, Version 1. (2023) doi:10.15497/RDA00099.
12. Lin, D. et al. The TRUST Principles for digital repositories. *Sci. Data* **7**, 144 (2020).
13. Caswell, I. Open peer review: what is it and what is UCL Press doing? UCL Press <https://uclpress.co.uk/open-peer-review-what-is-it-and-what-is-ucl-press-doing/> (2024).
14. Martinez-Ortiz, C. et al. Practical Guide to Software Management Plans. <https://zenodo.org/records/7589725> (2023) doi:10.5281/zenodo.7589725.
15. De Jong, M. & Scheers, M. Domeinarchitectuur Onderwijs & Flexibilisering. <https://www.surf.nl/files/2022-02/hosa-domeinarchitectuur-flexibilisering-onderwijs-v1.0.pdf> (2021).
16. Leijnse P. & Scheers, M. Domeinarchitectuur Identiteit en Toegang; <https://www.surf.nl/files/2023-01/hosa-domeinarchitectuur-iam-versie-1.0.pdf>
17. Van Veen, T. & Scheers, M. Domain Architecture for Research Data Management. <https://www.surf.nl/files/2023-06/hosa-domeinarchitectuur-research-data-management-v1.0eng-gb-1.pdf>.

A1. INLEIDING

De projectgroep heeft de OSI voor een groot deel ontwikkeld op basis van use-cases van workflows van onderzoekers en onderzoeksondersteunende professionals. Dit met het oog op het zoveel mogelijk bottom-up ontwikkelen van de OSI. De meeste use-cases waren gericht op het FAIR delen van onderzoeksresultaten, hetgeen de projectgroep ziet als de hoofdfunctie van OSI.

Samengevat omvat de methode de volgende stappen:

- Per use-case werd op basis van de stappen in de workflow de benodigde technische elementen geïdentificeerd en welke daarvan federatief zouden moeten zijn, de sociale en menselijke elementen die een rol spelen bij de stappen in de workflow en de federatieve afspraken die nodig zijn wat betreft standaarden en interoperabiliteit.
- In een tweede stap werden per domein de in de use-cases geïdentificeerde federatieve oplossingen in een federatief schema weergegeven met daarin de lijst van federatieve afspraken (technisch, semantisch, juridisch) en de door de federatieve oplossing te verbinden modules.
- De federatieve oplossingen zoals gepresenteerd in de federatieve schema's vormden de basis voor de voorstellen in hoofdstuk 3 voor de verschillende componenten van de OSI infrastructuur.



Figuur 8 Use-cases gekoppeld aan de onderzoekscyclus [LMA Research Data Management Working Group]

In Figuur 7 staan de use-case gekoppeld aan de onderzoekscyclus (de use-cases in geel zijn dubbel weergegeven omdat ze bij meerdere fasen van betekenis kunnen zijn). De use-cases zijn geplot bij die research fasen waar zij primair bij horen. Figuur 7 laat zien dat er weinig use-cases zijn uitgewerkt voor de uitvoering van onderzoek. De verwachting is dat dergelijke use-cases een sterk disciplinair karakter zullen hebben. In SPII Deel 2 dient bekeken te worden of er use-cases voor de

uitvoeringsfase door stakeholders aangedragen worden en in hoeverre hieruit additionele gewenste componenten voor de federatieve infrastructuur uit voortvloeien.

In de hierna volgende paragrafen worden de use-cases en de federatieve schema's gepresenteerd.

A2. OSC: USE-CASES EN FEDERATIEF SCHEMA

Voor OSC werden de volgende use-cases ontwikkeld:

- Publiceren van niet-traditionele onderzoeksresultaten
- Publiceren van een artikel in een traditioneel journal – in de huidige situatie en in de voorgestelde gewenste situatie
- Publiceren volgens PRC model
- Toegang tot wetenschappelijke publicaties voor een breed publiek

Deze use-cases worden op de volgende pagina's gepresenteerd.

Use-case: Publiceren van een rapport als voorbeeld van een niet-traditioneel onderzoeksresultaat	Technische componenten • Federatieve oplossing voor CRIS/repositories: mechanisme dat ervoor zorgt dat een rapport of andere output met meerdere auteurs van verschillende kennisinstellingen uitgewisseld wordt tussen de repositories/CRIS systemen op een zodanige manier dat er 1 DOI wordt toegekend, álle auteurs ORCID's krijgen en dat er maar 1 service organisatie metadateert (deduplicatie én efficiëntie)
Stappen van de workflow: Onderzoeker(s): • Finale versie opstellen • Goedkeuring opdrachtgever • DOI aanvragen bij institutioneel repository/CRIS service organisatie en in rapport opnemen • Licentie kiezen en in rapport opnemen • PDF maken en deponeren in repository/CRIS Repository/CRIS organisatie (bibliotheek): • Metadateren incl. check op PIDs ORCID, ROR, DOI, ...	Sociale/ menselijke elementen • Repository/CRIS service organisatie (meestal bibliotheek) biedt services aan 'als een uitgever' met aanvullingen van PIDs en metadatering in Marc21 format. • Repository/CRIS service organisaties van de universiteiten, hogescholen (PubliNova) en andere kennisinstellingen harmoniseren m.b.t. opname en metadatering beleid
Resultaat: publicatie en disseminatie • Rapport geregistreerd in CRIS • Rapport gepubliceerd in repository[§] • Metadata in bibliotheekcatalogus, automatische opname in WorldCat; ook in Google en andere zoekmachines	Standaarden en interoperabiliteit • PIDs: ORCID, ROR, DOI, in de toekomst mogelijk RAiD • Metadata standaard Marc21 • Dezelfde formats voor de verschillende typen output
Opmerkingen: • We maken geen onderscheid tussen CRIS en repository: deze zijn soms in één applicatie (Pure), soms in twee verschillende applicaties. (CRIS is een registratiesysteem, repository is een document opslag) Beiden vervullen een functie op het gebied van metadata-verzameling met koppeling naar opslag van objecten. • Hier is de workflow van de WUR weergegeven. Getoetst moet worden of dit overal vergelijkbaar is. • De federatieve oplossing dient technisch uitgewerkt te worden. De combinatie van OpenAIRE Provide, OpenAIRE Graph en OpenAIRE Broker kan mogelijk hiervoor dienen. Conclusie: • In deze use-case fungeert de institutionele repository als het publicatiekanaal naar de wereld.	

[§] Institutionele repositories bevatten voornamelijk tekstuele onderzoeksoutput. Het woord repository wordt ook gebruikt voor datarepositories, die onderzoeksdatasets omvatten.

PUBLICEREN ARTIKEL IN TRADITIONEEL JOURNAL – HUIDIGE SITUATIE

Use-case: Publiceren van een tijdschriftartikel [huidige situatie]	Technische componenten - Journal browser – hulp bij tijdschrift kiezen - Metadata via uitgever, via Scopus e.a.		
Onderzoeker(s): - Manuscript opstellen - Als tijdschrift: - Journal kiezen en daarin compliant zijn met eventuele subsidievoorwaarden en instellingsbeleid <i>Submit, review, rebuttal, acceptance</i> - Als PRC model: manuscript op preprint server, evt. later in (overlay) journal	Sociale/ menselijke elementen - CRIS/Repository service organisatie is niet/ nauwelijks <i>in the loop</i> tussen auteur en uitgever/preprint service en loopt daardoor achter de feiten aan - Vaak weinig motivatie bij onderzoeker m.b.t. vraag om full text van allang gepubliceerd artikel (voor onderzoeker bureaucratie en niet meer nuttig)		
CRIS/Repository organisatie (meestal bibliotheek):			
Als subscription journal	Als journal met Read & Publish deal of Publish deal	Als OA journal zonder Publish deal	
Metadata worden later ontvangen via Scopus o.i.d.	Bibliotheek krijgt bericht van acceptance via uitgever	Metadata worden later ontvangen via Scopus o.i.d.	
Vraagt aan auteur AAM voor opname in repository	Ontvangt metadata en published article voor opname in repository	Harvest published article voor opname in repository	
Resultaat - Gepubliceerd artikel in journal - Latere opname in repository en CRIS van de metadata (vaak via een levering van uitgever of Scopus o.i.d.) - Opname van full tekst of via (handmatige) downloads of via verzoek aan auteur voor AAM	Standaarden en interoperabiliteit - PIDs: ORCID, ROR, DOI, in de toekomst mogelijk RAiD - Metadata standaard Marc21		
Conclusie: CRIS/repository verzorgt een registratie en opslag ná de werkelijke publicatie.			

Use-case Publiceren tijdschriftartikel [gewenste situatie]			Technische componenten
Onderzoeker(s): - Manuscript opstellen - Deponeren in CRIS/ repository - Als tijdschrift: - Journal kiezen en daarin compliant zijn met eventuele subsidievoorwaarden en instellingsbeleid <i>Submit, review, rebuttal, acceptance</i> - Als PRC model: preprint server kiezen (zie hierna) CRIS/Repository organisatie (meestal bibliotheek): - Metadateren en PID-check (technische en administratieve controle) - Op basis van keuze auteurs via de federatieve oplossing ‘publicatiekanaal’ naar gewenste uitgever (of preprint server, zie hierna) - Indien niet aangesloten op de ‘publicatiekanaal’, dan gemetadateerd artikel via submit procedure door auteur naar uitgever			Journal browser – hulp bij tijdschrift kiezen Federatieve oplossing voor CRIS/repositories: mechanisme dat ervoor zorgt dat een rapport of andere output met meerdere auteurs van verschillende kennisinstellingen uitgewisseld wordt tussen de repositories/CRIS systemen op een zodanige manier dat er 1 DOI wordt toegekend, alle auteurs ORCID's krijgen en dat er maar 1 service organisatie metadateert (deduplicatie én efficiëntie) (zie OpenAIRE mechanismen hiervoor) Federatieve oplossing ‘publicatiekanaal’: het door de bibliotheek aangevulde manuscript wordt op basis van keuze door de auteur ingediend bij het gewenste publicatiekanaal. Mogelijk eerst te ontwikkelen voor preprint servers en voor uitgevers met Publish deals Federatieve oplossing ‘teruglevering published version’: zie publication router van Jisc
Subscription journal	Journal met Read & Publish deal of Publish deal	OA journal zonder Publish deal	
Als geaccepteerd, AAM en metadata al in CRIS/repository	Bibliotheek krijgt bericht van <i>acceptance</i> via uitgever Metadata al in CRIS/repository; AAM later vervangen door published versie	Metadata al in CRIS/repository Harvest published article voor opname in repository	Sociale/ menselijke elementen - CRIS/Repository service organisatie is in de loop en voegt waarde toe voor de auteur én uitgever/preprint service: - Metadata - PID check - Trust: manuscript komt van een betrouwbare kennisinstelling, is van bestaande auteur(s) bij die instelling en is traceerbaar (administratieve controle) - Toegevoegde waarde vanuit perspectief van onderzoekers en uitgevers nader te onderzoeken. - Vereist omslag in het denken van onderzoekers om hun concept-artikel eerst naar het repository te sturen voor aanvulling van metadata en PIDs en de gewenste publicatieroute aan te geven.
Teruglevering van published version aan CRIS/repository	Teruglevering van published version aan CRIS/repository	Teruglevering van published version aan CRIS/repository	

Resultaat: Gepubliceerd artikel in journal Metadata/AAM in vroege fase in repository.	Standaarden en interoperabiliteit • PIDs: ORCID, ROR, DOI, in de toekomst mogelijk RAiD • Metadata standaard Marc21
Conclusie: • In deze opzet wordt het CRIS/repository een eerste stap in het publiceren van tijdschriftartikelen. • Nader onderzocht dient te worden of dit voldoende toegevoegde waarde levert aan zowel onderzoekers als aan uitgevers • Dit vereist mogelijk een opschaling van de serviceorganisatie rond CRIS-systemen en repositories bij de instellingen (meer metadateringswerk) • Publicatiekanaal mogelijk eerst aan te sluiten op Diamond journals	

Use-case: Publiceren artikel volgens PRC model	Technische componenten:																		
Stappen van de workflow: Onderzoeker(s): - Onderzoeker biedt manuscript aan institutional repository - Bibliotheek verrijkt manuscript met metadata A. Met de federatieve oplossing ‘Publicatiekanaal’ wordt het manuscript vanuit het repository op basis van de keuze van de auteur doorgestuurd naar het gewenste publicatiekanaal. Dit waarborgt direct een zekere integriteit van het manuscript bij de preprintserver (is immers van een auteur van een gevestigde kennisinstelling). B. Als ‘publicatiekanaal’ niet is geïmplementeerd, dient de auteur zelf het manuscript in bij de preprint server <table><tr><td>ArXiv; MedRxiv; BioRxiv</td><td>Open Research Europe/ F1000 [of platform X-NL]</td><td>NL Aggregator publicatie-objecten</td></tr><tr><td>Postpub peer review Evt. aanpassen manuscript</td><td>Auteur zoekt zelf twee peer reviewers; mogelijk aanpassingen manuscript</td><td>Postpub peer review Evt. aanpassen manuscript</td></tr><tr><td>Indienen bij SCOAP3 jnls or andere journals</td><td></td><td>Via Notify indienen bij overlay journal</td></tr><tr><td>Editorial board/ Peer review</td><td></td><td>Editorial board/ Peer review</td></tr><tr><td>Publicatie peer-reviewed versie</td><td>Publicatie peer-reviewed versie</td><td>Publicatie peer-reviewed versie in repository</td></tr><tr><td>Connectie met repository (d.m.v. doorgeven DOI) en mechanisme teruglevering full tekst</td><td>Connectie met repository (d.m.v. doorgeven DOI) en mechanisme teruglevering full tekst</td><td>Connectie via Notify tussen overlay journal en repository</td></tr></table>	ArXiv; MedRxiv; BioRxiv	Open Research Europe/ F1000 [of platform X-NL]	NL Aggregator publicatie-objecten	Postpub peer review Evt. aanpassen manuscript	Auteur zoekt zelf twee peer reviewers; mogelijk aanpassingen manuscript	Postpub peer review Evt. aanpassen manuscript	Indienen bij SCOAP3 jnls or andere journals		Via Notify indienen bij overlay journal	Editorial board/ Peer review		Editorial board/ Peer review	Publicatie peer-reviewed versie	Publicatie peer-reviewed versie	Publicatie peer-reviewed versie in repository	Connectie met repository (d.m.v. doorgeven DOI) en mechanisme teruglevering full tekst	Connectie met repository (d.m.v. doorgeven DOI) en mechanisme teruglevering full tekst	Connectie via Notify tussen overlay journal en repository	Federatieve oplossing voor CRIS/repositories: mechanisme dat ervoor zorgt dat een rapport of andere output met meerdere auteurs van verschillende kennisinstellingen uitgewisseld wordt tussen de repositories/CRIS systemen op een zodanige manier dat er 1 DOI wordt toegekend, álle auteurs ORCIDs krijgen en dat er maar 1 service organisatie metadateert (deduplicatie én efficiëntie) Federatieve oplossing ‘publicatiekanaal’: het door de bibliotheek aangevulde manuscript wordt op basis van keuze door de auteur ingediend bij het gewenste publicatiekanaal. Mogelijk eerst te ontwikkelen voor preprint servers en voor uitgevers met Publish deals Federatieve oplossing Notify mechanisme: COAR’s Notify mechanisme werkt bij publicatie door overlay journals, uit te zoeken of dit ook kan werken tussen preprint server/ ORE platform en repository Federatieve oplossing ‘teruglevering published version’: voor publicatie in journals of in ORI, zie publication router van Jisc Platform X-NL: mogelijk in te richten platform voor Nederlandse publicaties met vergelijkbare functionaliteit van ORE. NL Aggregator publicatie-objecten: mogelijk in te richten platform voor Nederlandse publicaties met pre-print functionaliteit, mogelijk te combineren met X-NL
ArXiv; MedRxiv; BioRxiv	Open Research Europe/ F1000 [of platform X-NL]	NL Aggregator publicatie-objecten																	
Postpub peer review Evt. aanpassen manuscript	Auteur zoekt zelf twee peer reviewers; mogelijk aanpassingen manuscript	Postpub peer review Evt. aanpassen manuscript																	
Indienen bij SCOAP3 jnls or andere journals		Via Notify indienen bij overlay journal																	
Editorial board/ Peer review		Editorial board/ Peer review																	
Publicatie peer-reviewed versie	Publicatie peer-reviewed versie	Publicatie peer-reviewed versie in repository																	
Connectie met repository (d.m.v. doorgeven DOI) en mechanisme teruglevering full tekst	Connectie met repository (d.m.v. doorgeven DOI) en mechanisme teruglevering full tekst	Connectie via Notify tussen overlay journal en repository																	
	Sociale/ menselijke elementen - Publicatiecultuur verschillen tussen disciplines; hangt van de communities binnen disciplines - Time to publication kan incentive zijn - Mogelijke doelstelling: 10-15% van alle publicaties volgens PRC model.																		

Resultaat: • Publicatie van preprint en/of peer-reviewed artikel • In geval van federatieve oplossing ‘publicatiekanaal’: (1) manuscript is bereikt; (2) zekere waarborg van integriteit van het manuscript bij preprint server of publicatieplatform. • Teruglevering van full tekst aan repository door federatieve oplossing	Standaarden en interoperabiliteit • PIDs: ORCID, ROR, DOI, in de toekomst mogelijk RAiD • Metadata standaard Marc21
Opmerkingen: • De toegevoegde waarde van een NL Aggregator publicatie-objecten ligt in het volgende: (1) vanuit het <i>composability</i> perspectief is het makkelijker om publicatiekanalen te maken vanuit 1 platform naar preprintservers dan vanuit tientallen repositories; (2) de inrichting van het Notify mechanisme voor overlay journals is voor overlay journals alleen interessant bij voldoende volume; (3) als er ook een Postpub peer review optie wordt ingericht, dan kan ook niet-traditionele onderzoeksoutput worden gereviewd en dus erkend/gewaardeerd. • Peer review: ○ Postpub peer review: we bedoelen hier commentaar door lezers bij een gepubliceerde versie. ○ Peer review: we bedoelen hier het peer review proces, gecoördineerd door een journal editor of geïnitieerd door de auteur zelf (author-centered peer review). Wat betreft de openheid van peer review zijn er 7 kenmerken (zie o.a.¹³).	

Use-case: Gebruik en toegang tot wetenschappelijke informatie voor een breed publiek	Technische componenten: • Portal • Aggregatie van journalistieke artikelen • Aggregatie van wetenschappelijke artikelen • Wikipedia achtig publiceer systeem • AI tools
Stappen van de workflow: Gebruikers waarvoor geen vertaalslag nodig is (Findability en Accessible is knelpunt – goed overzicht van publieke bronnen (AI-tools erbij nemen zoals RAG): • Ca. 50% van wetenschappelijke artikelen is al open • Findability: Google Scholar, Dimensions, PubMed • Findability en Accessible: awareness raising van Open Alex, OpenAire: Als geen toegang (Closed Access), dan geen echte oplossing; turnaways tellen en dan vaststellen waar veel belangstelling voor is en dan actie daarop nemen. Gebruikers waarvoor vertaalslag nodig is: • Journalistieke vertaling: Groen Kennisnet, NEMO kennislink, wetenschapsbijlagen in kranten, ... • Directe toegang tot wetenschappelijke artikelen met vertaalslag: gezaghebbend portaal inrichten: hier is de informatie inclusief ondersteuning (bijv. AI-samenvatting van de top-10 van een zoekactie) • Over aantal veelgevraagde onderwerpen een ‘Wikipedia’ achtige pagina ontwikkelen door verschillende universiteiten – actieve voorlichting door wetenschappers: een kanaal voor wetenschappers voor publieksvoorlichting {wie is verantwoordelijk voor de kwaliteit}	Sociale/ menselijke elementen • Redactie • Commitment van universiteiten wetenschapsvoorlichting voor breed publiek • Meetellen in erkennen en waarderen
Resultaat: • Wetenschapsportal gericht op NL publiek	Standaarden en interoperabiliteit • Interoperabel met repositories en NL research portal • AI-bewerking altijd gekoppeld aan originele bron
Opmerkingen: • Het brede publiek ‘zit’ voornamelijk in de actualiteit, wetenschap heeft daar niet altijd antwoord op. Mogelijke oplossing: portal bouwen rond aantal actuele thema’s. • Mogelijk kan het portal beperkt worden tot een toegang tot open wetenschappelijke artikelen met ondersteuning in de vorm van AI-generated samenvattingen.	

Functies: Federated CRIS/ Institutional repositories en institutionele repositories als publicatiekanaal	
Federatieve oplossing(-en)	Afsprakenstelsel t.b.v. de federatie
Federated CRIS/institutionele repositories: Mechanisme dat ervoor zorgt dat een rapport of andere output met meerdere auteurs van verschillende kennisinstellingen uitgewisseld wordt tussen de repositories/CRIS systemen op een zodanige manier dat er 1 DOI wordt toegekend, alle auteurs ORCIDs krijgen en dat er maar 1 service organisatie metadateert (deduplicatie én efficiëntie). [bijv. mogelijk door combinatie van OpenAIRE PROVIDE, OpenAIRE Graph en OpenAIRE Broker]. Federatieve oplossing ‘teruglevering published version’: voor publicatie in journals of in ORE (mechanisme à la Publication router van Jisc).	Federated CRIS/institutionele repositories: • Harmonisering opname beleid CRIS/repositories • Afsprakenstelsel m.b.t. metadatering research output van auteurs van meerdere instellingen in CRIS/repositories: Bijvoorbeeld: instelling die eerst verantwoordelijk is, registreert output en stuurt een notificatie naar overige betrokken instellingen in NL. Betrokken instelling ontleent metagegevens. Archivering van objecten o.b.v DOI in één repository (verantwoordelijke instelling) • Harmonisering metadatabeleid en PID-beleid (zie ook hieronder) • Invoer ROR als instellingsidentificer op meerdere lagen (instelling/departement)
Institutionele repositories als publicatiekanaal: NL Aggregator publicatie-objecten met publicatiekanalen: (a) Aggregatie op landelijk niveau van publicatieobjecten vanuit de CRIS/repository systemen. Mogelijk aan te vullen met Postpub review optie. (b) Publicatiekanalen: het door de serviceorganisaties aangevulde manuscript wordt op basis van keuze door de auteur ingediend bij het gewenste publicatiekanaal. Mogelijk eerst te ontwikkelen voor preprint servers, voor het platform ORE, en voor uitgevers met Publish deals (c) Federatieve oplossing Notify mechanisme: COAR's Notify mechanisme werkend maken voor de aggregator op landelijk niveau	Institutionele repositories als publicatiekanaal: • Uitgangspunt zijn gefedereerde CRIS-systemen/institutionele repositories met bijbehorend afsprakenstelsel (zie hiervoor) • Harmonisering van het beleid van de serviceorganisaties (veelal de bibliotheken) bij de repositories en CRIS systemen, wat betreft opnamebeleid, metadata en PIDs, én timing van het verzorgen van de input. E.e.a. vereist mogelijk een uitbreiding van deze serviceorganisaties vanwege toegenomen volume en hogere eisen (vooral qua timing). • Omslag in denken van onderzoekers: het vereist een omslag in het denken van onderzoekers om hun concept-artikel eerst naar het repository te sturen voor aanvulling van metadata en PIDs en de gewenste publicatieroute aan te geven.
Door de federatie te verbinden modules	
Federated CRIS/institutionele repositories: • CRIS/institutionele repository systemen Institutionele repositories als publicatiekanaal: • Nationale bibliotheek eDepot (archiving/back-up) • DOI registries • ROR registries • Repository aggregator – IR's • Repository aggregator – OpenAire, Open indexing	

Toelichting:

We stellen twee stappen voor:

- **Optimalisatie en efficiëntieverbetering van de CRIS/repository systemen:** een federatie voor het invoeren van publicatie-objecten en de metadatering ervan (efficiëntie en ontdubbeling) en een vorm van de Jisc publication router om het harvesten van de full text zoveel mogelijk te automatiseren.
- **Inzetten van de CRIS/repositories infrastructuur als publicatiekanaal:** In het proces van outputregistratie → institutional repository → aggregatie, ligt de aandacht op registratie (achteraf) en het verzorgen van Open toegankelijkheid. Met deze infrastructuuronderdelen verzorgen de instellingen al een deel van het publicatieproces waarmee wetenschappelijke output vindbaar wordt gemaakt. Met name voor output die niet naar uitgevers gaat is dit proces al feitelijk de stap van intern document naar publieke beschikbaarheid zoals hierboven werd geschetst voor onderzoeksrapporten. In dit proces krijgen sommige documenttypen ook al een “institutional” branding. Deze “publicatiefunctie” kan veel krachtiger kan worden ingezet dan nu het geval is

Overige punten:

- Op termijn kan het doel zijn om CRIS-systemen vervangbaar te maken door federatieve oplossingen (zie SOLID project SURF met Universiteit van Gent)
- Efficiëntieslag, ontdubbeling van records en hogere dekkingsgraad full tekst in repositories zorgt voor hogere kwaliteit Green OA voor NL publicaties.
- Duidelijker scheiding van repository (open toegang tot objecten) en CRIS (output registratie op lokaal niveau door auteurs en support teams) is nodig om enerzijds proces minder redundant te maken en anderzijds meer te harmoniseren ten behoeve aggregatoren en eindgebruikers.
- Faciliteren van publiceren van niet-traditionele onderzoeksresultaten (incl. optie voor Postpub review); preprints, van publiceren via PRC model (overlay journals en mogelijk via ORE)
- Bijdragen aan een hogere kwaliteit van disseminatie (en dus zichtbaarheid).

A3. RDSM: USE-CASES EN FEDERATIEVE SCHEMA'S

Voor RDSM werden de volgende use-cases opgesteld:

- Delen van sensitieve data
- Hergebruik van sensitieve data
- Publiceren van single-use research software gerelateerd aan een publicatie
- Publiceren van herbruikbare research software

Deze worden in de pagina's hieronder gepresenteerd.

DELEN VAN SENSITIEVE DATA

Use-case delen van sensitieve dataset	Technische componenten
Stappen in de workflow: - Onderzoeker moet in staat worden gesteld om gevoelige data in een veilige omgeving te deponeren, inclusief informatie over eigenaarschap en wijze van eventuele toegang verkrijgen (data availability statement). - De mate van gevoeligheid van de data bepaalt daarbij het niveau van bescherming en toegangscontrole (dit impliceert dat criteria voor de bepaling hiervan – laag/medium/hoog/ extreem beschikbaar zijn) - Onderzoeker moet in staat worden gesteld om metadata m.b.t. de dataset te publiceren	- Data transfer tools, die de veiligheid van de transfer waarborgen (secure data transfer tools) - Rule-based access mechanismen [data brokers [DAB, RDX] - Encryptie data op opslaglocatie - Decryptie - Encryptie sleutel management - Back-up methods/schedules, system maintenance, testprocedures encryptie/decryptie - Multifactor Authenticatie/Toegangscontrole
	Sociale/ menselijke elementen - Eigenaarschap (en verantwoordelijkheden ingeregeld. Met name wanneer de onderzoeker niet meer werkzaam is bij de instelling. Wie is aanspreekpunt?) - Data transfer/sharing agreement (data access committee) - Beleid/procedures voor toegang, incl. ethische en juridische randvoorwaarden - Risicomanagement
Resultaat - Deponeren dataset (+ richtlijnen toegang, contextuele informatie, processing tools, consent informatie,..) - Citatiegegevens dataset - Publiceren metadata	Standaarden en interoperabiliteit - Licentie types - Metadata - Bestandsformaten
Toelichting: Het delen en hergebruiken van open datasets werkt in de praktijk reeds relatief goed. Het voornaamste knelpunt daarin wordt gevormd door de vaak gebrekkige metadata van de datasets die de vindbaarheid en bruikbaarheid kunnen beperken.	

HERGEBRUIK SENSITIEVE DATA VAN EERDER ONDERZOEK

Use-case hergebruik sensitieve data van eerder onderzoek	Technische componenten • Veilige omgeving voor data analyse (<i>compute to data</i> bij zeer gevoelige data [SANE] óf downloaden van data toegestaan) • Rule-based access mechanismen [data brokers [DAB, RDX] • Data transfer tools
Voorwaarden: • Data is gedeponneerd in een veilige omgeving • Inclusief bijbehorende licentie en verantwoordelijkheden (wie geeft toegang, wie krijgt toegang) en onder welke condities • Voldoende beschrijvende metadata Onderzoeker: • Na aanvraag dient de data-beherende instelling toestemming te verlenen [knelpunt: vaak is de betreffende onderzoeker niet meer werkzaam] • Toegang wordt verleend; encryptie sleutels overgedragen • Hergebruik en publicatie resultaten van het hergebruik • Citatie dataset	Sociale/ menselijke elementen • Duidelijke afspraken met de wie de data gedeeld mogen worden (toetsingskader kennisveiligheid) • Kwaliteit metadata en documentatie van de context • Data curatie
Resultaat • Hergebruik dataset • Citatie dataset	Standaarden en interoperabiliteit • PIDs om dataset te kunnen vinden en te kunnen citeren • File formats • Date documentatie • Licenties en contracten • Context (hoe zijn de data gegenereerd, welke technische apparatuur is gebruikt; kan een andere onderzoeker de data begrijpen)
Voor het vinden van datasets zijn er vier opties: • Search engines (zoals DataCite Search (van DataCite), Dataset Search (van Google) en Mendeley Data (Elsevier) • Publication aggregators met een filter voor datasets (zoals Dimensions.ai, Lens.org, OpenAire) • Harversters (zoals ODISSEI) • Data repositories (zoals DANS, 4TU.ResearchData)	

PUBLICEREN SINGLE-USE RESEARCH SOFTWARE GERELATEERD AAN EEN PUBLICATIE

Publiceren single-use research software gerelateerd aan een publicatie	Technische componenten
Stappen van de workflow	• Archiefoplossing (bijv. 4TU.ResearchData of Zenodo of Software Heritage); version control systeem • Registry voor de vindbaarheid: Research Software Directory • PID resolvers
• Script/ code beschrijven d.m.v. documentatie • Deponeren script/code en documentatie waarbij in elk geval nodig is: metadata met een link naar de dataset, een link naar de publicatie (evt. een link naar de project identifier) en licentie. • Mogelijk: registreren bij Research Software Directory (bij deponeren in 4TU.ResearchData gebeurt dit automatisch, RSD harvest ook metadata van Zenodo).	• Sociale/ menselijke elementen • Kwaliteit van de documentatie • Curatie van de documentatie (versiebeheer) • Recognition op beschikbaarstelling (aan de hand van <i>contribution guidelines</i> indien meerdere software engineers) • SBOM – Software Bill of Materials (nog in ontwikkeling; lijst van aspecten van open source software met. o.a. dependencies, veiligheidsrisico's en zwakke plekken).
Resultaat	Standaarden en interoperabiliteit
• Vindbaar en herbruikbaar research softwarepakket (t.b.v. reproduceerbaarheid)	• PIDs (voor software DOI en/of SHHID; verder DOI voor publicaties en datasets; ORCID voor onderzoekers en software engineers; RAiD voor projecten) • Metadata standaarden • API standaarden • licenties
• De RAiD identifier en de SPDX identifier en de metadatastandaarden CFF en Codemeta (exchange scheme for software metadata) bestaan reeds, maar dienen breder geadopteerd te worden. In technische zin is alles aanwezig voor de archivering van software met het oog op linken aan publicaties. In SPII Deel 2 zou e.e.a. ontwikkeld kunnen worden tot een algemeen aanvaard 'Research Software Management Framework' teneinde de adoptie ervan te stimuleren. • Research Software Directory is een linking pin tussen software publicaties, data, projecten, mensen en organisatie.	

PUBLICEREN VAN HERBRUIKBARE RESEARCH SOFTWARE

Publiceren herbruikbare research software	Technische componenten
Stappen van de workflow • Version control platform voor de source code (GitHub/GitLab). • Kiezen van een licentie (zie bv. TLDRlegal) • Toepassen van een versioning scheme zoals semantic versioning. • Citeerbaar maken van de software door de relevante metadata toe te voegen d.m.v. bijvoorbeeld CFF • Waarborgen software kwaliteit (bv. codestyle, static analysis, tests, CI infrastructuur) • Maken van releases en deze archiveren met een eigen identifier (bv. een DOI van Zenodo). Hiernaast ook een DOI voor de verzameling van alle versies (concept DOI op Zenodo). • Beschikbaar maken documentatie • Publiceren van de software op de relevante registries, zoals de Research Software Directory, bio.tools, ASCL.net, etc. • Evt. ook publiceren van <i>contribution guidelines</i>, <i>issue tracking</i> en governance, SBOM, etc. • Registreren bij Research Software Directory (bij deponeren in 4TU.ResearchData gebeurt dit automatisch, RSD harvest ook metadata van Zenodo; vaak dienen de metadata nog te worden aangevuld).	<i>In computeromgeving van softwareontwikkelaar:</i> • Version control systeem • Testing/CI infrastructure (zoals linters, unit and integration test frameworks, en de hardware om deze te draaien). <i>(inter)nationaal niveau:</i> • Archiefoplossing (bijv. 4TU.ResearchData of Zenodo) • Registry voor de vindbaarheid: Research Software Directory • PID resolvers
Resultaat • Vindbaar, herbruikbaar en citeerbaar research softwarepakket • Verbeterde reproduceerbaarheid	Sociale/ menselijke elementen • Kwaliteit van de documentatie, tests, etc. • Curatie van de documentatie (versiebeheer) • Recognition op beschikbaarstelling (aan de hand van <i>contribution guidelines</i> indien meerdere software engineers, gebruik van citatie metadata) • SBOM – Software Bill of Materials (nog in ontwikkeling; lijst van aspecten van open source software met. o.a. <i>dependencies</i>, veiligheidsrisico's en zwakke plekken. • Het linken van de research software met gerelateerde research output en activiteiten (publicaties, data, projecten) evenals mensen en organisaties. Dit is wat de Research Software Directory aanbiedt. Standaarden en interoperabiliteit • PIDs (voor software DOI, SWID; verder DOI voor publicaties en datasets; ORCID voor onderzoekers en software engineers; RAiD voor projecten) • Metadata standaarden zoals CFF voor citatie metadata en Codemeta voor software beschrijvingen. • API standaarden • Licenties (d.m.v. SPDX identifier – <i>standardized abbreviations for common open source software licenses</i>)
Toelichting: • De RAiD identifier en de SPDX identifier en de metadatastandaarden CFF en Codemeta (exchange scheme for software metadata) bestaan reeds, maar dienen breder geadopteerd te worden. In technische zin is alles aanwezig voor de archivering van software met het oog op linken aan publicaties. In SPII Deel 2 zou e.e.a. ontwikkeld kunnen worden tot een algemeen aanvaard 'Research Software Management Framework' teneinde de adoptie ervan te stimuleren. • Research Software Directory is een linking pin tussen software publicaties, data, projecten, mensen en organisatie.	

Functie: delen en hergebruik (sensitieve) data	
Federatieve oplossing(-en) • Data management omgevingen (zoals Yoda) • Data opslag (opslag tijdens het onderzoek; vaak integraal gekoppeld aan een data management omgeving. En al dan niet gemeenschappelijk) • Data repositories (Yoda biedt bijv. een archiveringsoptie, maar zonder functionaliteit voor long-term preservation) • Long-term data repositories (met long-term preservation functionaliteit, zoals DANS, 4TU.ResearchData) • Data transfer tools, die de veiligheid van de transfer waarborgen (secure data transfer tools) • Rule-based access mechanismen [data brokers [zoals DAB, RDX] • Toegangsbeheermechanismen: toegang tot data én data management omgeving voor individuele onderzoekers waarmee wordt samengewerkt (bijv. SRAM (SURF Research Access Management); multi-factor-authenticatie • Veilige omgeving voor data analyse (<i>compute to data</i> bij zeer gevoelige data [zoals SANE]^h)	Afsprakenstelsel t.b.v. de federatie • Harmoniseren criteria sensitieve data • Afspraken over encryptie/decryptie procedures • Afspraken over verantwoordelijkheden t.a.v. toegang en data governance • Afspraken over data transfer/sharing agreements • Afspraken over risicomanagement
Door de federatie te verbinden modules • Data management omgevingen en data repositories • Data management omgevingen en (lange termijn) data opslag omgevingen • Data management omgevingen en data analyse omgevingen • Verbindingen met de research administratie omgeving	
Toelichting: • We onderscheiden 8 federatieve componenten. Sommige data management omgevingen bieden meerdere componenten ineen. Bijv. Yoda biedt data management, opslag en archivering. • Verbindingen (ketenintegratie) tussen de researchadministratie met o.m. data management plannen enerzijds en het onderzoeksproces anderzijds op een manier die de onderzoeker faciliteert zijn wenselijk. • Dataopslag kan federatief maar ook per instelling georganiseerd te worden. Data opslag kan worden gezien als onderdeel van research infrastructuur.	

^h <https://www.surf.nl/themas/onderzoeksinfrastructuur/sane-veilige-omgeving-voor-analyse-van-gevoelige-data>

Functie: delen en hergebruik research software	
Federatieve oplossing(-en) • Research Software Directory • PID resolvers	Afsprakenstelsel t.b.v. de federatie • PIDs voor software (zoals DOIs of SWID) • SPDX identifier voor licenties • metadatastandaarden CFF en Codemeta (exchange scheme for software metadata) • Gebruik software management plan¹⁴
Door de federatie te verbinden modules • Archieven als Zenodo, Software heritage, 4TU.ResearchData • Version control platformen zoals Github, GitLab, etc.	
Toelichting: • De DOI/SWID identifiers en de SPDX identifier en de metadatastandaarden CFF en Codemeta (exchange scheme for software metadata) bestaan reeds, maar dienen breder geadopteerd te worden. In technische zin is alles aanwezig voor de archivering van software met het oog op linken aan publicaties. In SPII Deel 2 zou e.e.a. ontwikkeld kunnen worden tot een algemeen aanvaard 'Research Software Management Framework' teneinde de adoptie ervan te stimuleren. • Wat betreft het samenwerken en gebruiken van open source research software wordt gebruik gemaakt van Github of van GitLab. Daarbij wordt het volgende opgemerkt: Github wordt nu door research software ontwikkelaars veelal gebruikt met privéaccounts, hetgeen leidt tot versnippering en verwarring waar wat staat. GitLab heeft naast een algemene site ook lokale implementaties bij sommige instellingen, die niet toegankelijk zijn voor anderen.	

A4. ORI: USE-CASES EN FEDERATIEF SCHEMA

Voor ORI werden er twee use-cases ontwikkeld:

- evaluatie van een onderzoeksgroep
- analyse van onderzoeklandschap met het oog op nieuwe inzichten

Deze use-cases worden hieronder gepresenteerd, gevolgd door een federatief schema.

EVALUATIE VAN EEN ONDERZOEKSGROEP

Use-case Evaluate a research group	Technical components
Steps of the workflow • Define research and staff participation in group • What projects did they participate in and for how many FTE? • List per project: ○ FTE participants ○ Grant info ○ External collaboration partners ○ Research output (including patents) ○ Equipment • Output: ○ Publications: #, FAIR score; Impact research: normalized citation score; Impact society: media ○ Software: #, FAIR score, Github stars & forks ○ Data: #, FAIR score, volume; reuse/downloads ○ Patents: #, royalties €; # licence takers ○ Collaboration partners: # academic, # industry/government; countries/global ○ Research equipment: # usage, value	• CRIS and institutional systems (project management system, HR system, financial system) • Funders' database (Cordis, NWO, ...) • Citation database • Altmetrics database
	Sociale/ human elements • SEP protocol • Research intelligence support • Valorisation office to select and validate research output • Linking research output to projects (now missing) and data on collaborations (also missing)
Result • Overview production • Insight in impact • Insight in value for money	Standards en interoperability • Metadata standards for output, persons, projects • PIDs: DOI, ORCID. RAiD, GrantID
Toelichting: • Voor impact van de research output worden externe systemen gebruikt (bijvoorbeeld Open Alex en CrossRef voor citaties). Meestal zijn deze getallen een proxy voor de werkelijke impact, die moeilijk objectief meetbaar is.	

ANALYSE VAN ONDERZOEKSLANDSCHAP MET HET OOG OP NIEUWE INZICHTEN

Use-case Analyse research landscape for new insights [example: benchmark for a specific topic].	Technical components (example) Altmetrics, Pubmed, Databricks, PowerBI
Steps of the workflow ▪ Select scope (which institutions, which topics) ▪ Create classifier/apply it to publications ▪ Create dataset selections and pull in additional information sources ▪ Build dashboard	Sociale/ human elements Research intelligence team
Result ▪ <u>dashboard</u>	Standards en interoperability DOI, ROR, MESH (or other vocabulary)
Toelichting: Zie twee voorbeelden: https://app.powerbi.com/view?r=eyJrIjoieYjdMOTJkYTETZWNmMi00NWU0LTlkYzgtNjhkOTBkMzc5YWE0IiwidCI6IjQ2MmE5YzljLTJkYTETNGJmYi1iMzE4LTBmODMwNDNmZTQ5YiIsImMiOiJh9 https://epdos.nl/nl/rarediseasemonitor/	

Open Research Information	
Federatieve oplossing(-en) • Metadata hub: collects metadata <i>and</i> connects related metadata	Afsprakenstelsel t.b.v. de federatie • Information exchange between Federated modules using OpenAIRE CERIF ○ Currently OpenAIRE CERIF 1.2.0 ○ Extend to Dutch requirements; agreement via EduStandaard.nl • Making connections using PIDS: ROR, ORCID, DOI, RAiD, Grant ID, SWHID, RRID ISSN, ISBN, ISNI, WikiData, EPIC, ARK
Door de federatie te verbinden modules: Research input entities: organisational units, persons, grants, projects, data management plans, equipment [Within NL-OSI: to be collected for research administration platform, CRIS systems] [Outside NL-OSI: to be collected from among others NWO API, CrossRef, FundRef & Grant ID, RAID] Research output entities: publications, datasets, software, patent, events (e.g. media performances) [Within NL-OSI: to be collected from the CRIS systems; data/software/institutional repositories, publication platforms] [Outside NL-OSI: to be collected from oa CrossRef, DataCite,...] Research outcome entities: views, downloads, commits, stars, forks, citations (scientific; reference in peer-reviewed article), mentions (societal; reference in news article, policy doc, etc) [Within NL-OSI: views & downloads to be collected from the CRIS systems; data/software/institutional repositories, publication platforms] [Outside NL-OSI: views & downloads to be collected from publishers (article level COUNTER reports) and OpenAIRE; commits, stars & forks on external Git platforms (e.g. Github, Codeberg, Git lab etc); citations from Open Citations; mentions from Crossref Event tracker.] Enrichment of Research entities: Modules that text/data mine full text of article for funder information, SDG's, other labels etc. [Within OSI code algorithms used for inferencing, textmining, etc. are shared in the Software module]	
Toelichting: OpenAIRE currently collects and connects federated research input, output and outcome entities and performs enrichments from various sources, within and outside the NL-OSI (OSI = the proposed Open Science Infrastructure)	

USE-CASE RESEARCH ADMINISTRATIE EN COMPLIANCE

Research administratie en compliance met regelgeving	Technische componenten • Administratief platform waarin verschillende componenten geïntegreerd beschikbaar zijn (zoals een data management plan tool, ethische toetsing, AVG-registratie) • Per instelling een opslagomgeving voor administratieve gegevens m.b.t. onderzoeksprojecten • Bij voorkeur een koppeling met andere administratieve omgevingen van de instelling (finance, HR, project control, CRIS)
Stappen in de workflow: • Onderzoeker moet in staat worden gesteld om met minimale tijdsinvestering en zo min mogelijk dubbel werk alle administratieve afhandeling van zijn/haar onderzoeksproject m.b.t. data management te kunnen doen • Dit betreft o.m. data management (voor, tijdens en na het onderzoek), software managementplan, Data Protection Impact Assessment (DPIA), Ethics review, contracten met bedrijven voor, tijdens en na onderzoek. Daarnaast is het verweven met links naar adviseurs in de instellingen over deze onderwerpen. Er is een workflow aspect in opgenomen dat de onderzoeker hierdoor heen leidt. • Administratieve gegevens m.b.t. onderzoeksproject zijn controleerbaar en opgeslagen • Inclusief data licentie en verantwoordelijkheden (wie geeft toegang, wie krijgt toegang) en onder welke condities en wie is verantwoordelijk voor verwijderen/langer bewaren van data (data governance).	Sociale/ menselijke elementen • Een support desk en/of data support organisatie • Duidelijke afspraken over welke administratieve handelingen vereist zijn
Resultaat • Kwalitatief hoogwaardige administratieve gegevens m.b.t. onderzoeksproject • Documentatie onderzoeksdata is beschikbaar en bruikbaar • Directe koppeling met support	Standaarden en interoperabiliteit • PIDs om onderzoeksprojecten te identificeren (RAiD, Grant-ID) • Data documentatie • Licenties en contracten
Toelichting: • Dit betreft niet uitsluitend Open Science. Niettemin is het van groot belang voor Open Science in termen van transparantie en verifieerbaarheid en bepaalt het in hoge mate in hoeverre het onderzoeksproject en de resultaten daarvan open, open onder condities of gesloten kunnen zijn. Vandaar dat het opgenomen in OSI. • Volgens onze informatie zijn er tenminste vijf instellingen bezig met het ontwikkelen van een systeem hiervoor. Dit leidt tot onnodige fragmentatie, terwijl een gezamenlijk systeem recht kan doen aan instellingspecifiek beleid. • Bij het administratieve platform denken we aan een gedelegeerd/multitenant systeem dat SURF zou kunnen beheren en dus met voor elke instelling een eigen 'instance'. • De invoering van de project identifier RAiD is essentieel, evenals de Grant-ID voor projectfinanciering.	

Functie: research administration/ compliance met wet- en regelgeving	
Federatieve oplossing(-en) • Administratief platform (multitenant opzet) waarin verschillende componenten geïntegreerd beschikbaar zijn (zoals een data management plan tool, ethische toetsing, DPIA) dat gekoppeld is per instance aan de relevante administratieve omgevingen van de betreffende instelling • Opslagomgeving (institutioneel) voor: ○ Administratieve gegevens m.b.t. onderzoeksproject (daardoor controleerbaar) ○ Ook opslag van data licentie en verantwoordelijkheden (wie geeft toegang, wie krijgt toegang), toegangscondities en wie is verantwoordelijk voor verwijderen/langer bewaren van data (data governance).	Afsprakenstelsel t.b.v. de federatie • Invoering RAiD identifier (Research Activity Identifier) • Afspraken over metadata en datamodellen (bijvoorbeeld machine-actionable DMP's) • Support desk
Door de federatie te verbinden modules • Instances van het multi-tenantsysteem automatisch met elkaar verbonden	
Toelichting: • Vanuit het Open Science perspectief komt hieruit naar voren: (1) belang van ketenintegratie 'meenemen' van tijdens de verschillende stappen gegenereerde metadata die relevant zijn voor delen, hergebruik en reproduceerbaarheid en (2) automatisering van de data access mechanismen waardoor onderzoekers ondersteund worden om datasets met toegang onder condities te delen.	

BIJLAGE B: TOELICHTING SCOPE SPII DEEL 1

B1. INLEIDING

In paragraaf 1.2 is de scope van SPII Deel 1 puntsgewijs weergegeven. In deze bijlage wordt e.e.a. toelicht.

B2. OPEN SCIENCE INFRASTRUCTUUR AFGEBAKEND VAN RESEARCH INFRASTRUCTUUR:

SCOPE

Het is de bedoeling om een infrastructuurplan te ontwikkelen voor Open Science en niet voor een digitale research infrastructuur. zoals bijvoorbeeld High Performance Computing, GWI's etc.). Anders gezegd: OSI is het Open Science perspectief op de (digitale) research infrastructuur. Daarbij gaat het om welke aanpassingen en aanvullingen de digitale research infrastructuur nodig heeft om Open Science te faciliteren. Aansluiting op de hierboven genoemde research infrastructuren komt in Deel 2 aan bod.

TOELICHTING

Research infrastructuur omvat digitale middelen die nodig zijn om onderzoek uit te voeren, waaronder laboratoriumapparatuur, High Performance Computing, tools, en dergelijke. De focus ligt op het ondersteunen van het onderzoeksproces en is in veel gevallen gericht op discipline-specifieke onderzoeksprocessen.

OSI kan gezien worden als die functionaliteit van de digitale research infrastructuur, die ingericht is om onderzoek en onderzoeksresultaten maximaal open en herbruikbaar te maken, terwijl de overige functionaliteit van de digitale research infrastructuur gericht is op digitale middelen en diensten die de andere aspecten van het onderzoek – open of gesloten – ondersteunen. OSI gaat dus om aanpassingen of aanvullingen op de research infrastructuur die Open Science mogelijk maken, waarbij het einddoel is dat alle research infrastructuur functionaliteit biedt die Open Science faciliteert.

B3. FEDERATIEVE INFRASTRUCTUUR ALS UITGANGSPUNT

SCOPE

- Het uitgangspunt is SPII zoveel mogelijk te baseren op wat er al is aan bestaande systemen. Daarom stelt SPII een federatieve/federatie-achtige opzet voor OSI voor, modulair opgebouwd en integraal van opzet. Een te ontwikkelen referentiearchitectuur voor OSI (à la HOSA) maakt deel uit van deze opzet.
- SPII Deel 1 focust op de technische vormgeving. De voor een federatieve structuur benodigde governance, organisatie, middelen en commitment vormen deel van het voorgestelde SPII Deel 2.

FEDERATIEVE EN FEDERATIE-ACHTIGE MODELLEN

We kunnen de volgende modellen onderscheiden die relevant kunnen zijn voor de Open Science infrastructuur:

- **Centraal model:** Alle software, processen en controles zijn ondergebracht bij een centrale instantie/server/dienst.
- **Aggregatie model:** Data worden periodiek opgehaald (geaggregeerd) uit verschillende bronnen en samengevoegd in een centrale databank of index.
- **Gedelegeerd model/multitenant model:** Een gedeeld softwaresysteem wordt door verschillende instellingen gebruikt, maar met eigen configuraties. Elke 'tenant' van het systeem draait een eigen variant.
- **Federatief model:** Een netwerk van autonome partijen die samenwerken op basis van afspraken en standaarden: er is geen centrale opslag of controle maar onderlinge interoperabiliteit.

Het aggregatiemodel en het gedelegeerd/multitenant model kunnen beschouwd worden als 'federatie-achtig'. Omwille van de leesbaarheid zijn in dit rapport ook de federatie-achtige vormen aangeduid met federatief.

In de tabel hieronder staan ter illustratie enkele kenmerken en voorbeelden van deze modellen weergegeven.

Modellen infrastructuur	Centraal model	Federatie-achtig	Federatie-achtig	Federatief model
		Gedelegeerd/ multitenant model	Aggregatie model	
Opzet	Eén centraal systeem	Eén gedeelde implementatie met lokale 'instances'	Meerdere systemen, metadata centraal samengevoegd	Verschillende systemen met gemeenschappelijke standaarden
Eigen data per instelling	Mogelijk, maar centraal opgeslagen	In aparte instances	Lokaal opgeslagen en centraal samengevoegd	Volledig onder lokaal beheer
Autonoom beleid mogelijk	Nee	Deels zelfstandig	Bij bron	Autonoom m.u.v. inachtneming federatieve afspraken
Voorbeelden	Klassieke database, nationaal register	CRIS-systeem met lokale invulling zoals PURE; DSpace Yoda, Dataverse.NL	OpenAIRE (d.m.v. harvesting) Unpaywall (d.m.v. crawling) ODISSEI portal	Federatief identiteitsbeheer, EOSC

Tabel 1 Vergelijking modellen voor digitale infrastructuur

ONDERZOEKSINFRASTRUCTUUR EN HORA/HOSA

De Hoger Onderwijs Referentie Architectuur (HORA) en Hoger Onderwijs Sector Architectuur (HOSA) kunnen gezien worden als een implementatie van de Nederlandse Overheid Referentie Architectuur (NORA) voor onderzoek en onderwijs.

De HOSA is een sectorarchitectuur en kent drie vertalingen naar ‘domeinarchitecturen’: Onderwijs¹⁵, IAM¹⁶ en RDM¹⁷.

De digitale onderwijs-Infrastructuren zijn tamelijk centraal georganiseerd (in vergelijking met de onderzoeksinfrastructuur) en de HOSA domeinarchitectuur voor onderwijs beschrijft dan ook een redelijk gecentraliseerde, platformgerichte architectuur, die momenteel stapsgewijs geïmplementeerd wordt. De HOSA domeinarchitectuur voor IAM beschrijft een ontwikkeling van een meer centrale architectuur naar een federatieve opzet en voorziet een verdere ontwikkeling naar een meer decentrale structuur.

Onderzoeksinfrastructuren zijn gefragmenteerd en omvatten zowel instellingspecifieke en nationale componenten als discipline-specifieke, veelal internationale componenten. Een gecentraliseerde infrastructuur voor onderzoek bestaande uit een of enkele platformen is dan ook niet aan de orde.

B4. RDSM, OSC EN ORI

SCOPE

SPII focust op een OSI infrastructuur voor Research Data & Software Management, Open Scholarly Communication infrastructuur en Open Research Information en op het aspect FAIR delen van onderzoeksresultaten daarbinnen. SPII verwacht daarmee een belangrijk deel van de behoefte aan OSI te dekken.

TOELICHTING

De focus voor SPII Deel 1 ligt op Research Data & Software Management (RDSM), Open Scholarly Communication infrastructuur (OSC) en Open Research Information (ORI) infrastructuur. Het doel daarbij is om het creëren, delen en hergebruiken van FAIR research data en software, Open Access publicaties en Open Access tot andere research output, zoals tools, protocollen, methoden en open-source hardware te ondersteunen en om open research informatie over onderzoek aan de Nederlandse instellingen te verbeteren.

Met de focus op RDSM, OCS en ORI wordt het overgrote deel van de infrastructuur voor Open Science gecoverd. Infrastructurele voorzieningen voor andere onderdelen van Open Science, zoals bijvoorbeeld voor Citizen Science, kunnen dan in een later stadium van de ontwikkeling van OSI ontwikkeld en gekoppeld worden.

B5. DISCIPLINE-AGNOSTISCH IN DEEL 1

SCOPE

SPII Deel 1 richt zich op OSI-functionaliteiten die relevant zijn voor Nederlandse onderzoeksprojecten ongeacht hun discipline. De focus ligt daarbij in eerste instantie op gegevens die de instellingen met elkaar kunnen en willen delen op een effectieve en efficiënte manier door middel van een federatieve infrastructuur. De aansluiting op meer op disciplines gerichte Open Science infrastructurele componenten en op disciplinaire research infrastructures komt in Deel 2 aan bod.

TOELICHTING

Generieke Open Science infrastructurele componenten betreffen generieke modules zoals algemene data repositories, software repositories, CRIS-systemen, institutionele repositories, preprint servers, publicatieplatforms, en persistent identifier systemen (zoals ORCID, DOI en ROR). Daarnaast zijn er meer op specifieke disciplines gerichte Open Science infrastructurele componenten, zoals domeinspecifieke data repositories, vakgebiedsgerichte preprintservers en publicatieplatformen, en open-source software tools voor specifieke zaken als de analyse van *genomic data* uit lab experimenten of software voor astronomen.

BIJLAGE C: VERTREKPUNTEN VOOR DE ONTWIKKELING VAN DE OSI REFERENTIEARCHITECTUUR

C1. DERA EN HET NDE PROGRAMMA ALS VOORBEELD

Het programma [Netwerk Digitaal Erfgoed](#) bestaat 10 jaar en opereert in een divers landschap van erfgoedinstellingen. De kern van het programma kan samengevat worden als het verbinden van digitaal erfgoed. Vanaf het begin van het programma is er gewerkt aan een referentiearchitectuur – [DERA](#) – een afgeleide van de NORA referentiearchitectuur.

Men is gestart met een eerste versie van DERA waarin een aantal principes werd geformuleerd. In de loop van het programma is DERA telkens uitgebreid en bijgewerkt. De respondent geeft aan dat een dergelijke referentiearchitectuur nooit is uitontwikkeld.

De DERA fungeert als een soort anker en leidt soms tot een vertaling naar concrete requirements aan bijvoorbeeld collectiebeheersystemen. Deze zijn gebaseerd op DERA en ook opgesteld in wisselwerking met de leveranciers van deze systemen (wat is realistisch en wat is haalbaar). De resultaten daarvan zijn weer ingebracht in DERA. Inmiddels zijn deze requirements aanvaard en kunnen instellingen DERA-compliant collectiebeheersystemen aanschaffen

De DERA is dus soms leidend, maar soms ook volgend op ontwikkelingen in het veld. Deze ontwikkelingen worden gemonitord en besproken in een periodiek overleg over de DERA en indien relevant ingebracht in een nieuwe versie.

De conclusie luidt dat DERA een sturende en harmoniserende werking heeft gehad op de ontwikkelingen rond digitaal erfgoed. Het fungeert als een stip aan de horizon, die richting geeft aan wat er in de praktijk gebeurt. Als belangrijkste *lesson learned* noemt de respondent de scope van de referentiearchitectuur: het is essentieel om de doelen ervan steeds in het oog te houden, zodat de complexiteit ervan niet onnodig groot wordt.

C2. VERTREKPUNTEN OSI REFERENTIEARCHITECTUUR

Gedurende het project SPII Deel 1 zijn uit de diverse besprekingen een aantal vertrekpunten afgeleid, die op de volgende pagina in tabelvorm zijn weergegeven. Deze tabel dient gezien te worden als een bottom-up verzameling van observaties, ideeën en best-practices welke nog nadere bespreking en verfijning behoeven. Enkele belangrijke hoofdlijnen uit de tabel zijn:

- Maak het eenvoudig voor de onderzoeker
- Creëer metadata bij de bron
- Ontwikkel een geïntegreerd metadatamodel voor alle vormen van onderzoeksinformatie
- Ondersteun zowel conventionele als innovatieve publicatiemodellen
- Zorg voor een open opzet bij de ontwikkeling van OSI.

Principle	Subprinciple	Rationale	Consequences
Put the researcher experience front and center	- Make it easy and efficient to capture as much information as possible - create value for the researcher - enter only once		
	- Put into place a national repository that accepts preprints, published articles, theses, reports and other documents. - Offer institutional entry points to the repository. Allow relevant support staff to review the submissions to be adequate and properly described. Expose metadata through APIs and embed JSON-LD to ensure indexing. - Verify authors and their consent to publication at least through email and ORCID validation.	- Make it easy for anyone who is affiliated with a member institution to submit a document. Support entry of complete and high-quality metadata with automated extraction (potentially AI based), autocompletion, database lookups and PID support.	
Enable both conventional and innovative publication models	- Establish a submission platform that supports configurable workflows for journal articles and books. - Integrate with the national repository, other repositories and preprint servers. - Consider a generic peer review platform that can be used for solicited or unsolicited review in any kind of publication model.		- Items to be reviewed can include publications, datasets and software. Promote open peer review and provide metadata about reviews. All reviewers need to be authenticated. Allow other sites (e.g. repositories and journals) to pull in reviews and display them.
	- Center publication flows around the document. Instead of sending around files, leave the publication in place and add layers of annotations to it. - Support different versions of publications and allow for living, evolving documents.		

Principle	Subprinciple	Rationale	Consequences
Collect research information along the research lifecycle	- Connect publications as much as possible with each other and with other outputs such as datasets and software. - Integrate with an archive for automatic preservation of items.		
	- Choose a project-based view on the research lifecycle to integrate administrative workflows and their associated metadata - Use RAiD as identifier - Include data/software management plans, ethical reviews, GDPR documentation/information security classification, AI impact assessments, and other required documentation	- much information about research is generated along the way but not captured and only reconstructed afterwards - reporting becomes much easier	- Start working groups with the respective experts from institutions to harmonize these documents as much as possible and create machine-readable versions - Lobby NWO to use this metadata as reporting instead of manual entry in ISAAC
Curate metadata in source systems	- Start producing high quality metadata at the source instead of investing a lot of time into curating, merging, fixing data afterwards. - Do less data manipulations in the CRIS - Curate metadata about research outputs in the respective repositories - Aggregate metadata from source systems (like repositories) and let CRISs subscribe to that. Do not aggregate metadata from CRISs as it is not always clear where it originally comes from. - Expose all metadata so it can be aggregated	- information that is curated in downstream systems is difficult to extract, leads to duplicate efforts and to conflicting metadata - the provenance of metadata in non-source systems is not always clear	
	- Put into place a system that allows institutions to model their internal organization and to record affiliations of researchers. Allow institutions (and researchers) to claim an output for a specific person/organization affiliation. Base reporting on this.		

Principle	Subprinciple	Rationale	Consequences
An integrated, extensible, straightforward metadata model for all research information	- Choose existing work as basis, e.g. schema.org with JSON-LD serialization - Push for use of linked data in all systems. - Make a distinction between entities (publications, persons, organizations etc.) and records about them.	- a linked data model allows for extension, missing information and connection between data sources - schema.org is widely adopted and provides a very rich vocabulary for research specific metadata - This allows for explicit modelling of provenance, uncertainties in data quality, and integration of data sources.	- create adapters for systems not using the integrated metadata model yet - Develop the metadata model fully in the open on GitHub through pull requests and allow adoption by anyone
	- Offer a cohesive data storage experience that makes it easy to move data into different storage solutions throughout the research lifecycle, and to a repository for publishing. - Make it possible to automatically check the stored data and software against the relevant data and software management plans. - Aggregate metadata for datasets and software and make it discoverable through search, recommendation and query interfaces.		- Enabled collaboration on the data and mounting of data sources in analysis environments.
	- Consider using a single platform for collaborative software development. Integrate it with Software Heritage and the Research Software Directory. Make it easy to pull code from the platform into the analysis environments.		
Develop Open Science infrastructure openly	- Discuss standards and software development publicly. - Encourage participation from anyone interested and adhering to a code of conduct.	- open discussion that documents the thought process through which a decision was reached - inclusive participation is a principle of Open Science and leads to solutions with wider support	

Principle	Subprinciple	Rationale	Consequences
Benefit from work done in other communities	- Take the spirit of open-source development to build infrastructure.	- Open-source practices are well established and have led to impressive outcomes - Learning, sharing and reuse are at the center of Open Science	

BIJLAGE D: SUGGESTIES VOOR DE START VAN SPII DEEL 2 EN 3

INLEIDING

In de laatste dagen van het project SPII Deel 1 zijn nog een aantal opmerkingen en suggesties naar voren gekomen die van belang kunnen zijn voor de voorgestelde projecten SPII Deel 2 en Deel 3.

SCOPE OPNIEUW BEZIEN

Er is een voorstel binnengekomen om met een aantal experts op het gebied van Citizen Science aan de hand van enkele use-cases op dat gebied te bezien in hoeverre de voorgestelde OSI infrastructuur de infrastructurele behoeften van Citizen Science dekt.

Ook is op basis van andere opmerkingen is meer aandacht aan digitale samenwerkingsomgevingen in Deel 2 op zijn plaats.

SPLITSING DEEL 2 EN DEEL 3

Het risico van voorgestelde splitsing is dat de resultaten van Deel 2 stuiten op een gebrek aan organisatie en financiering om deze daadwerkelijk uit te voeren. Dit met het oog op een eerder project over een mogelijke Open Knowledge Base, waar veel stakeholders tijd in hebben gestoken maar wat desondanks niet van de grond gekomen is. Een dergelijke situatie zou mogelijk ondervangen kunnen worden door meerdere scenario's wat betreft gezamenlijke organisatie en financiering op te stellen en deze al in een eerder stadium bestuurlijk te bespreken.

VERKENNING OM USE-CASES MOGELIJK UIT TE BREIDEN

Er zijn meerdere suggesties genoteerd om het aantal use-cases uit te breiden en te bezien of dit aanleiding geeft tot aanpassingen in de federatieve schema's en mogelijk tot aanpassingen of aanvullingen voor de OSI infrastructuur. Deze suggesties behelzen onder meer:

- Use-cases met betrekking tot de uitvoeringsfase van onderzoeksprojecten
- Use-cases die alle domeinen integreren
- Use-cases vanuit een onderzoeksinstituut of onderzoeksgroep
- Use-cases vanuit specifieke disciplines.

ALIGNMENT MET REFERENTIEARCHITECTUUR EN BASISDIENSTEN VAN EOSC EN NFDI

EOSC ontwikkelt een referentie architectuur en basisdiensten, die toegang, interoperabiliteit en betrouwbaarheid mogelijk maken. Het gaat onder meer om PID services, metadata registries (standaarden en vocabulaires, FAIR sharing, RDA registries), monitoring en accounting (gebruik, beschikbaarheid, prestaties van diensten) en *resource allocation*. De [NFDI](#) ontwikkelt ook een referentiearchitectuur en heeft vergelijkbare basisdiensten opgezet of is deze aan het opzetten. SPII Deel 2 doet er goed aan deze referentiearchitectuur en basisdiensten te bestuderen en waar nodig in lijn te brengen met de OSI voorstellen.

DOORKIJK TOEKOMST

SPII Deel 1 is uitgegaan van bestaande workflows. Het is echter duidelijk dat er technische ontwikkelingen plaatsvinden zoals Artificial Intelligence en decentralisatie, die mogelijk grote impact kunnen hebben op workflows van onderzoekers en onderzoeksondersteunende professionals en/of op de Open Science infrastructuur. Een exercitie die met deze ontwikkelingen rekening houdt, kan mogelijk een doorkijk opleveren voor de langere termijn en de deliverables van SPII Deel 2 toekomstbestendiger maken.

INVENTARISATIE RELEVANTE LOPENDE PROJECTEN

Aanbevolen wordt om te inventariseren welke projecten op het gebied van data infrastructuren, data interoperabiliteit, publicatieplatformen e.d. er momenteel lopen, waarvan de resultaten naar verwachting relevant kunnen zijn voor OSI.

OVERIGE SUGGESTIES

- **Referentiearchitectuur:** Als uitgangspunt zou ook ‘Open by design/FAIR by design’ meegenomen kunnen worden. Het gaat dan om het FAIR maken van de onderliggende infrastructuur.
- **Digilabs:** Het verdient aanbeveling om een verbindingen te leggen met een tweetal Digilabs. Het [Digilab](#) van de overheid werkt aan de technische interoperabiliteit en verantwoord delen van data. In een ander [Digilab](#) werken vijf TO2-organisaties en zes Rijkskennisinstellingen samen aan de ontwikkeling van een federatieve digitale infrastructuur.
- **GORC-model:** In de ontwikkeling van het [GORC-model](#) is veel denkracht ingezet. Het SPII Deel 2 project kan mogelijk gebruik maken van (een deel van) de resultaten van dit model.

Stuurgroep	
Naam	Organisatie
Darco Jansen	UNL
John Doove	SURF
Frans Oort	Chiefs Open Science
Sarah Coombs	Vereniging Hogescholen

Adviesgroep	
Naam	Organisatie
Jacko Koster	UNL
Dennis Eijsten	KB
Joost Elshoff	VH, DCC-PO, SHB
Madeleine de Smaele	4TUResearchData
Tom Bakker	eScience Center
Ellen Carbo	ZonMw
Hans de Jonge	NWO/ OSNL
Ella Bosch	NWO-I
Aline Korterink	KNAW
Niels Bolding	Health-RI
Lieuwe Kool	NFU
Bolinda Hoeksema	SHB

Projectgroep	
Naam	Organisatie
Till Bey	SURF
Maurice Vanderfeesten	SURF
Jason Maassen	eScience Center
Marcel Ras	UNL-VU
Patris van Boxel	DANS
Hubert Krekels	UNL-WUR
Claudia de Visser	SURF
Maurits van der Graaf	Pleiade

Geconsulteerd over versie 0.7		
Naam	Organisatie	Onderwerp
Ane van der Leij	RUG	Fit-gap OSC
Alenka Prinic	UL	Fit-gap OSC
Heleen Palmen	UL	Fit-gap OSC
Dirk van Gorp (RU)	RU	Fit-gap OSC
Jeroen Bosman (UU)	UU	Fit-gap OSC
Pascal Braak (uvA)	UvA	Fit-gap OSC
Alastair Dunning	TUD	Fit-gap OSC

Chris Baars	DANS	Fit-gap RDSM
Cees Hof	DANS	Fit-gap RDSM
Kees den Heijer	DCC-IN	Fit-gap RDSM
Leon van Ekeren	HBO-PO	Fit-gap RDSM
Maarten Hoogerwerf	SURF	Fit-gap RDSM
Celia van Gelder	TDCC-LSH	RDSM
Joanne Yeomans	TDCC-NES	RDSM
Nicole Emmenegger	TDCC-SSH	RDSM
Arjan Schalken	VU	Fit-gap ORI
Nees Jan van Eck	CWTS	Fit-gap ORI
Cristina Huidiu	WUR	Fit-gap ORI
Ludo Waltman	CWTS	Fit-gap ORI
Nick Feenstra	RUG	Fit-gap ORI
Abeer Pervaiz	VU	Research administratie
Margreet Riphagen	Hogeschool Utrecht	Research administratie
Mark Thompson	VU	IT architectuur
Taco Horsley	UvA	IT architectuur
Joyce Nijkamp	WUR	IT architectuur
Martijn Kleppe	KB	Algemeen
Hilde van Wijngaarden	VU	Algemeen
Anja Smit	DANS	EOSC-NL

BIJLAGE F: GLOSSARY

AAI	Authenticatie- en autorisatie-infrastructuur
Data Access Broken	Treedt op als een "makelaar" die onderzoekers verbindt met aanbieders van researchdatasets [voer een DOI in, indien open directe toegang, indien beperkte toegang doorverwijzing naar de landingspagina van de data provider.
Data transfer service	Een data transfer service is een opslagruimte waar data tijdelijk worden bewaard en verwerkt voordat deze worden verplaatst naar hun eindbestemming.
DataverseNL	DataverseNL is een research datarepository die wordt beheerd door DANS met afzonderlijke instances die door universiteiten/hogescholen worden beheerd.
DCC-IN	Nationaal implementatienetwerk van Digital Competence Centers (DCC-IN). Het DCC-IN verenigt IT-afdelingen en universiteitsbibliotheken van alle Nederlandse universiteiten, universitaire medische centra en KNAW- en NWO-instituten.
iRODS	iRODS is open-source datamanagement software die een scala aan onderzoeksdatamanagement (RDM) functionaliteiten biedt, zoals het veilig delen van data, opslagvirtualisatie, metadatamanagement, toegangscontrole en regelgebaseerde workflows. iRODS is vrij complexe middleware die verdere configuratie en integratie nodig heeft voordat het gebruikt kan worden door onderzoekers.
Nederlands Research portal	Aggregator van Nederlandse onderzoeksinformatie, gebaseerd op OpenAire
Odissei	ODISSEI (Open Data Infrastructure for Social Science and Economic Innovations) is de nationale onderzoeksinfrastructuur voor de Nederlandse sociale wetenschappen met diverse componenten, waaronder toegang tot data, expertise en computers.
RAiD	Research Activity Identifier (RAiD) is een persistent identifier (PID) en wereldwijd register voor onderzoeksprojecten. Gereguleerd door ISO-standaard
RDX	Research Data Exchange: (1) onderzoeker stelt dataset beschikbaar onder voorwaarden door een paar vragen te beantwoorden; (2) potentiële gebruiker wil toegang tot een dataset en tekent digitaal een contract waarmee hij/zij zich bindt aan die voorwaarden. [Dus de Data Access Broker helpt je toegang te krijgen tot de dataset, RDX automatiseert het proces van restrictieve toegang].
SANE	Werkruimte voor gevoelige gegevens (met 'compute to data')
Surf Onderzoekscloud	SURF Research Cloud is een gefedereerde omgeving/portaal waar je een virtuele onderzoeksomgeving kunt bouwen. Je kunt vooraf geconfigureerde workspaces en datasets gebruiken of ze zelf toevoegen. Voordelen: (1) reproduceerbare onderzoeksomgevingen creëren; (2) gebruik maken van de beste faciliteiten en (3) veilig samenwerken. Er is computing en er is storage.
TDCC	Thematische digitale competentiecentra: Life Science & Health (LSH); Natuur- en ingenieurswetenschappen (NES); Sociale en menswetenschappen (SSH)

URN:NBN-oplosser	De Uniform Resource Name (URN) is een Persistent Identifier (PID) waarmee - in tegenstelling tot URL's - online bronnen uniek en permanent kunnen worden geïdentificeerd, ongeacht waar ze zijn opgeslagen. nbn: Nationaal Bibliografisch Nummer
YODA	YODA is een applicatie voor instellingen die RDM ondersteunt gedurende de gehele onderzoekscyclus: van het veilig en eenvoudig opslaan en delen van gegevens tijdens het onderzoeksproces, tot het delen van gegevens binnen onderzoeksgroepen en -projecten en uiteindelijk tot het archiveren en publiceren van onderzoeksgegevens. YODA maakt gebruik van iRODS als basistechnologie. • Gebruiksvriendelijk (webgebaseerde, grafische gebruikersinterface en de WebDAV-interface) • Ondersteunt het hele onderzoeksproces van gegevensverwerving tot publicatie • Ondersteunt gestandaardiseerde workflows voor de opslag en archivering van onderzoeksgegevens. • Directe koppeling tussen gegevens en metadata (inclusief flexibele metadataschema's) • Datapublicatie en de bijbehorende aanmaak van Digital Object Identifiers (DOI's) • Een flexibele integratie met iRODS die ook geschikt is voor (zeer) data-intensief onderzoek.