

**Verrijkte
publicaties:
hoe verder?**

Colofon

Verrijkte publicaties: hoe verder?

SURFfoundation
PO Box 2290
NL-3500 GG Utrecht
T + 31 30 234 66 00
F + 31 30 233 29 60

info@surf.nl
www.surf.nl

Auteur

Martin Feijen

Eindredactie

Keith Russell (SURFfoundation)

SURF is de ICT-samenwerkingsorganisatie van het hoger onderwijs en onderzoek (www.surf.nl).
Deze publicatie is digitaal beschikbaar via de website van SURFfoundation:
www.surfoundation.nl/publicaties

© Stichting SURF
Februari 2010
ISBN 9789078887096

Deze publicatie verschijnt onder de Creative Commons licentie Naamsvermelding 3.0 Nederland.
<http://creativecommons.org/licenses/by/3.0/nl/>



Inhoudsopgave

1	Samenvatting	5
1.1	De conclusies	5
1.2	Aanbevelingen	5
2	Management summary	7
2.1	Conclusions	7
2.2	Recommendations	7
3	Inleiding	9
3.1	Opdrachtomschrijving	9
3.2	Werkwijze	9
3.3	Leeswijzer	10
3.3.1	Structuur	10
3.3.2	Begrippen	10
4	Huidige situatie: Stand van zaken in Nederland	11
4.1	Repositories	11
4.2	Diensten	11
4.3	Archieven	12
4.4	Organisatie	12
4.5	De tenderprojecten 2009	13
4.6	Andere nationale ontwikkelingen	15
4.6.1	SURFfoundation	15
4.6.2	Nederlandse Vereniging voor Beroepsbeoefenaren in de Bibliotheek-, Informatie- en Kennissector (NVB)	15
4.6.3	Rijksoverheid	15
5	Huidige situatie: Internationale ontwikkelingen	17
6	Vraag en aanbod	19
6.1	Vanuit de onderzoeker bekeken	19
7	Conclusies	21
7.1	Pluspunten	21
7.2	Belemmeringen	21
8	Aanbevelingen	23
	Bijlage 1 - Bronnen	27
	Bijlage 2 - Open data	29
	Bijlage 3 - Overzicht internationale ontwikkelingen	31
	Bijlage 4 - Overzicht repositories	37
	Bijlage 5 - De workshop verrijkte publicaties 25 november 2009	39
	Bijlage 6 - Verslagen van de gesprekken met tenderprojectleiders	43

1 Samenvatting

In dit rapport wordt een overzicht gegeven van de stand van zaken rondom het thema Verrijkte Publicaties per eind 2009. Het rapport is in opdracht van SURFfoundation geschreven als voorbereiding op het maken van plannen voor de komende twee jaar in het SURFshare Programma.

1.1 De conclusies

Pluspunten

De infrastructuur in Nederland is – gezien vanuit het perspectief van VPs – redelijk ver ontwikkeld. Er zijn al repositories en data-archieven, er zijn aanbieders van diensten, de organisatiegraad is vrij hoog – zeker in vergelijking met het buitenland – en de tenderprojecten hebben resultaten en ervaringen opgeleverd. Er is een kleine, maar sterke voorhoede, er is redelijk veel materiaal om mee te werken, er zijn tools en er is een relatief gunstig klimaat om in te werken.

Belemmeringen

De repository systemen die nu in Nederland gebruikt worden, moeten eerst nog geschikt gemaakt worden voor verwerking van VPs. Bij de instellingen is sprake van een vrij groot gat tussen voorhoede en achterhoede. Om verder te kunnen is het noodzakelijk dit gat te dichtten. Afbakening en concretisering van het begrip VP is nodig. VPs kosten nog veel tijd en zijn discipline specifiek. Het aanbod dient te worden verbonden met de vraag. Er is een (deels) latente behoefte bij onderzoekers aan ondersteuning. De onderzoeker praat dan niet over een VP maar over diensten die helpen bij de uitvoering van het onderzoekswerk. Om daarop in te spelen is een vertaalslag nodig van de interne terminologie en aanpak. Infrastructureel denken is nodig maar moet vertaald worden naar dienstgericht denken.

1.2 Aanbevelingen

Zorg voor beleidsontwikkeling op het thema academic services, zowel bij SURFfoundation als bij de SURFshare instellingen (i.c. universiteitsbibliotheken en HBO mediatheken). Leiderschap en visie vanuit SURFfoundation zijn onontbeerlijk. Zorg voor convergentie (focus) en gerichte aandacht op datgene wat de komende twee jaar nodig is om SURFshare succesvol af te ronden. Begin met eenvoudige VPs en streef naar een situatie waarbij eind 2011 alle HO instellingen eenvoudige VPs als onderdeel van hun academic services kunnen aanbieden. Alloceer de resources binnen SURFshare naar rato van hun belang voor de ontwikkeling van deze services. Voorkom versnippering. De koers voor de komende twee jaar moet zo snel mogelijk worden vastgesteld en daarna gecommuniceerd worden naar de SURFshare community.

Maak de bestaande repository systemen geschikt voor verwerking van VPs en ondersteun dit met gerichte aandacht vanuit SURFshare. Richt de aandacht daarna op de ontwikkeling van diensten onder het adagium dat kwaliteit belangrijker is dan kwantiteit.

Marketing en communicatie zijn noodzakelijk. Ontmoet de onderzoeker op zijn terrein. Start met eenvoudige diensten gebaseerd op eenvoudige VPs. Denk aan diensten voor opslag van onderzoeksdata, voor identificatie van die data, van artikelen en van personen, en voor het vergroten van zichtbaarheid voor de onderzoeker.

De diensten zouden vorm kunnen krijgen door een project in de komende twee jaar, waarin naast een centrale landelijke dienst ook lokale diensten worden ontwikkeld die specifiek inspelen op disciplines op lokaal niveau. De landelijke dienst is belangrijk voor de uniformiteit in infrastructuur en als basis voor de lokale diensten. De aanpak kenmerkt zich daarbij vooral door het werken met bestaande middelen en organisatie. Het gaat primair om het benutten van aanwezig potentieel in de aanwezige infrastructuur en niet om het ontwikkelen van nieuwe technologie. Via deze aanpak

kan het SURFshare Programma eind 2011 tot aansprekende resultaten leiden die inspiratie geven voor een vervolg.

2 Management summary

This report offers an overview of the current state of affairs of Enhanced Publications (EPs). It has been written for SURFfoundation in preparation of their planning of the SURFshare Programme for the coming two years.

2.1 Conclusions

Pluses

From the perspective of EPs we can state that the Dutch infrastructure is reasonably well developed. Repositories are in place, as well as operational data archives, service providers are active, the level of organisation is high and the tender projects have resulted in tools and experiences. A small but powerful group of early adopters is established, a broad range of materials to work with are available, there are tools and there is a good (political) climate to work in.

Barriers

Most repository platforms are not (yet) able to process EPs and a large gap looms between early adopters and followers. Additional work to align the infrastructure is required before further steps are possible. Demarcation of the conceptual and practical implications of EPs is necessary, the theme is too broad and too complex. The processing of EPs is time consuming and the process is discipline specific.

The services on offer need to be aligned with demand. Researchers do require support, but their needs are not (yet) clearly expressed. A researcher does not talk about an Enhanced Publication but about services that will help him do his work. The terminology used in the SURFshare programme needs to be translated into a language that is understandable to the research communities. We think in infrastructure but we need to talk about services.

2.2 Recommendations

It would be worth while to develop a clear policy for academic services, within SURFfoundation, as well as within SURFshare organisations (University Libraries and Academy Colleges). SURFfoundation should take on a leading role in this. Convergence of activities and focused attention is required for those items that are crucial for a successful completion of the SURFshare Programme in the next two years. An approach would be to start with "simple" EPs and set a goal to create the situation by the end of 2011 in which all organisations in Dutch Higher Education can support simple EPs as part of their academic services. SURFshare resources should be allocated according to their importance for development of these services. Fragmentation should be avoided. A timely decision should be made on the planning for the next two years and this should be communicated to the SURFshare community.

Ensure that existing repository systems can process EPs and support the required efforts through the SURFshare Programme. Next, focus on the development of services, keeping in mind that quality is more important than quantity.

This approach requires marketing and communication. Meet the researcher on his own territory. Start off with simple services based on simple EPs. Researchers are asking for services which will support the storage of research data, the identification of articles, data and persons, and improve the visibility of their work.

Development of the services could take place through a two year project focused on developing a central national service and specific local services aimed at disciplines and groups at the local level. The national service is important as this can ensure a uniform infrastructure and will form the foundation upon which local services can be provided. The basic approach should make use of

existing resources and existing organisations. It is of paramount importance to develop potential within the existing infrastructure, the development of new technology itself is not the goal. With this approach, the SURFshare programme can bring inspiring results by the end of 2011 that may lead to further development in this field.

3 Inleiding

In de eerste twee jaar van het SURFshare programma is op een breed terrein een groot aantal activiteiten uitgevoerd: verrijkte publicaties, collaboratories, auteursrechten, infrastructuur, CRIS, open access, HBO kennisdisseminatie en permanente toegang tot onderzoeksdata. In 2008 is SURFfoundation gestart met een reeks [tenderprojecten](#) op het gebied van Verrijkte Publicaties (hierna verder afgekort als "VPs"). Aan het einde van 2009 zullen in totaal zeven tenderprojecten zijn afgerond en zijn resultaten en ervaringen beschikbaar. Daarmee komt de vraag naar voren op welke wijze de tweede helft van de looptijd van het SURFshare programma voor dit thema ingevuld moet gaan worden. Zijn de resultaten van de tenders opschaalbaar? Is de organisatie en de infrastructuur ver genoeg ontwikkeld? Hoe verhoudt het thema verrijkte publicaties zich tot de andere SURFshare activiteiten? Wat willen de SURFshare instellingen zelf in de komende twee jaar op dit thema gaan doen, wat vindt men wel en wat vindt men niet belangrijk? Is het wenselijk om de aanpak van de eerste twee jaar SURFshare (divergentie) te wijzigen (convergentie) in de volgende twee jaar?

3.1 Opdrachtomschrijving

De opdrachtomschrijving vanuit het Platform Onderzoek voor het opstellen van dit rapport luidde als volgt:

Geef een analyse van de huidige stand van zaken voor het thema Verrijkte Publicaties en doe aanbevelingen voor volgende te nemen stappen op het gebied van de te ontwikkelen infrastructuur voor VPs. Hoe is de stand van zaken met betrekking tot VPs; welke elementen missen we nog; welke stappen zouden moeten worden gezet om in 2010 de basis-infrastructuur en workflow op orde te hebben om VPs te kunnen creëren, toegankelijk te maken, op te slaan en te communiceren. Kijk hierbij naar zowel de organisatorische kant van een infrastructuur voor VP's als naar de meer technische kant. Bovenstaande vragen moeten zijn ingekaderd binnen de meer algemene internationale ontwikkelingen die van belang zijn voor VP's, zoals o.a. het semantisch web en ontwikkelingen op het terrein van Linked Data.

3.2 Werkwijze

Bij het opstellen van dit rapport is de volgende werkwijze gehanteerd. De verschillende onderdelen van een infrastructuur voor VPs zijn in hun onderlinge samenhang bekeken:

- afspraken en richtlijnen over de gezamenlijke toepassing van standaarden
- afspraken over taakverdeling, rolverdeling, workflows
- organisatie
- software en systemen voor verwerking van VPs

Er zijn gesprekken gevoerd met de projectleiders van de tenderprojecten op het terrein van VPs: Escape, Dataplus, JALC, Proefschriften Plus en Veteran Tapes. De verslagen van deze gesprekken zijn toegevoegd als bijlagen bij dit rapport.

De analyse van de huidige stand van zaken in Nederland voor wat betreft infrastructuur en organisatie vond plaats op basis van gesprekken met SURFshare medewerkers, Koninklijke Bibliotheek, medewerkers van universiteiten en desk research. Tevens werd desk research gedaan naar de stand van zaken rond datasets en VPs in Nederland en internationaal.

De voorlopige resultaten van het onderzoek zijn gepresenteerd en besproken met vertegenwoordigers van de SURFshare community in een workshop eind november 2009. De resultaten van de workshop zijn verwerkt in deze rapportage. Eerdere versies van dit rapport zijn van commentaar voorzien door John Doove, Gerard van Westrienen en Leo Plugge. Waarvoor dank.

3.3 Leeswijzer

3.3.1 Structuur

Het vierde hoofdstuk geeft een tour d'horizon van het thema VPs, opgedeeld over drie onderwerpen. Hierbij beschrijft de eerste paragraaf de situatie in Nederland. In dit eerste onderdeel komen ook de zeven tenderprojecten die op het terrein van VPs zijn uitgevoerd aan de orde.

Hoofdstuk vijf geeft een beeld van de internationale situatie.

Het daarop volgende hoofdstuk gaat nader in op de (deels) latente behoefte van onderzoekers aan ondersteuning en geeft daarmee een schets van vraag en aanbod rond VPs als onderdeel van een breder dienstenpakket.

De conclusies worden beschreven in hoofdstuk zeven en de aanbevelingen zijn te vinden in hoofdstuk acht.

3.3.2 Begrippen

Het rapport is primair geschreven voor de leden van de SURFshare community. Voor de lezer die (nog) niet helemaal thuis is in het onderwerp Verrijkte Publicaties volgt hierna een kort overzicht van de voornaamste kernbegrippen.

Verrijkte Publicatie: een publicatie bestaande uit een artikel met daarbij behorende onderzoeksdata, modellen, figuren, statistieken enz. In het Engels spreekt men van een Enhanced Publication. Andere termen met ongeveer dezelfde betekenis: samengesteld document en compound object.

PID: persistent identifier, een aan een object of publicatie toegekend uniek nummer dat dient als unieke identificatie zonder dat het nummer wijzigt. Vergelijkbaar met ISBN voor boeken.

DAI: digital author identification, een aan een persoon toegekend uniek nummer dat dient als unieke identificatie van die persoon. In de context van dit rapport gaat het om DAI's die zijn toegekend aan onderzoekers verbonden aan een Nederlandse HO instelling

OAI - ORE: Open Archives Initiative : Object Reuse and Exchange. “..defines standards for the description and exchange of aggregations of Web resources. These aggregations, sometimes called compound digital objects, may combine distributed resources with multiple media types including text, images, data, and video. The goal of these standards is to expose the rich content in these aggregations to applications that support authoring, deposit, exchange, visualization, reuse, and preservation..” (ontleend aan: <http://www.openarchives.org/ore/>)

DIDL / MODS / IMS-CP: afkortingen die gebruikt worden ter aanduiding van standaarden die zijn ontwikkeld voor geautomatiseerd transport van VPs van het ene systeem naar het andere. In Nederland gebruiken we DIDL / MODS voor VPs en IMS-CP voor leermiddelen.

4 Huidige situatie: Stand van zaken in Nederland

In dit hoofdstuk kijken we naar de volgende aspecten:

- repositories
- diensten
- archieven
- organisatie
- tenderprojecten 2009
- andere ontwikkelingen

4.1 Repositories

Repositories zijn in Nederland in gebruik sinds 2003 en zijn inmiddels een vast onderdeel van de lokale voorzieningen in Nederlandse HO instellingen. Bijlage 4 geeft een gedetailleerd overzicht van de operationele repositories in het Nederlandse HO veld. Het aantal systemen dat wordt gebruikt is bijna net zo groot als het aantal instellingen. Het beeld is erg versnipperd. Dat wordt versterkt door het feit dat er drie verschillende versies van afspraken voor transport van metadata in gebruik zijn. Versie 1 van deze afspraken is vooral in gebruik bij HBO Kennisbank. Versie 2 is vooral in gebruik bij universiteiten, terwijl een klein deel daarvan (4 à 5 instellingen) versie 3 heeft geïnstalleerd. Alle universiteiten hebben zich gecommitteerd om begin 2010 versie 3 in gebruik te hebben. Deze versie is een voorwaarde voor effectief datatransport van VPs. Versie 3 ondersteunt o.a. DIDL / MODS, DAI, en een solide PID oplossing.

De meeste Nederlandse repositories zijn nog niet ingericht op de verwerking van VPs. Een platform als Fedora is daar wel geschikt voor, maar dat wordt door de meeste universiteiten niet gebruikt. In Utrecht wordt Dspace via lokale aanvullingen geschikt gemaakt voor verwerking van verrijkte proefschriften, maar dit is voornamelijk geen standaard Dspace oplossing die ook elders toegepast kan worden. In internationaal verband wordt gewerkt aan het geschikt maken van Dspace voor compound objects, zoals bijv. in het [JISC project Dspace 2.0](#)

Learning Object Repository systemen zijn wel ingericht op de verwerking van compound documents veelal op basis van de IMS-CP standaard. Deze wijkt af van de ORE standaard die binnen de SURFshare tenderprojecten wordt gebruikt.

4.2 Diensten

De drie belangrijkste diensten in Nederland zijn: Narcis, Lorenet en HBO Kennisbank.

[NARCIS](#) geeft toegang tot 350.000 publicaties, waarvan 200.000 open access publicaties, 9.100 datasets en daarnaast informatie over onderzoekers (expertise), onderzoeksprojecten en onderzoeksinstellingen in Nederland (status medio december 2009). Narcis zal volgens de huidige afspraken begin 2010 geschikt zijn voor de ontvangst van VPs vanuit systemen bij HO instellingen die voldoen aan de daarvoor afgesproken standaard (versie 3.0).

[Lorenet](#) is een portal voor leermiddelen, ontwikkeld in het SURF project Lorenet. De portal bevat circa 125.000 leermiddelen van twaalf Nederlandse HO instellingen of projecten (stand medio december 2009). Qua technologie en standaarden kan de portal als voorbeeld gebruikt worden voor VPs, hoewel de gebruikte standaard (IMS-CP) specifiek voor leermiddelen is ontwikkeld.

[HBO-Kennisbank](#) is een portal waarmee 13.500 scripties, artikelen en leermiddelen van negentien HBO instellingen ontsloten worden (stand medio december 2009).

Voor Lorenet en HBO Kennisbank zijn er voor zover bekend geen plannen voor verdere uitbouw van functionaliteit die relevant is voor het thema VPs.

4.3 Archieven

Naast de drie service providers zijn er vier archieven die een belangrijke rol spelen in de infrastructuur.

Organisatie	Discipline	Materiaal
KB E-Depot	Alfa, beta, gamma	Publicaties
DANS	Humaniora en sociale wetenschappen	Datasets
3TU	Techniek	Datasets
Beeld & Geluid	Audiovisueel	Datasets
Buitenlandse organisaties	Divers	Datasets

In de tabel zijn ook buitenlandse archieven benoemd omdat voor sommige vakgebieden (bijv. sociologie) in Nederland gemaakte datasets zijn opgeslagen in bijv. Duitse archieven.

Er zijn disciplines waarvoor nog geen algemene, openbare archieven zijn opgezet, zoals bijv. de medische wetenschappen. In het kader van dit rapport is hiervoor geen verder onderzoek gedaan. Het totale aantal datasets in deze archieven is op dit moment niet exact bekend. DANS bewaart circa 6500 datasets (waarvan een zeer groot deel uit de archeologie), het [3TU archief](#) omvat "...4 'collecties' van zeer verschillende omvang, sommige strekken zicht uit over enkele jaren, meerdere locaties en sensoren tot de resultaten van een enkele enquête. Twee collecties zijn in bewerking en er zijn gesprekken gaande met onderzoeksgroepen..."¹

Via Narcis worden meer dan 9000 datasets aangeboden. We mogen op basis van deze getallen aannemen dat het totale aantal in Nederlandse archieven aanwezige datasets boven de 10.000 ligt, met een jaarlijkse groei van ca. 2500 (best guess). We mogen verder aannemen dat het totale aantal datasets in Nederland veel hoger (een factor vijf) ligt, zowel qua beschikbaar volume als qua groei.

De KB heeft zich vooral toegelegd op de long term preservation van publicaties en heeft in het [DRIVER-II rapport](#) de problematiek rondom long term preservation van VPs uitstekend in kaart gebracht. Het huidige beleid van de KB is nog steeds primair gericht op het E-Depot voor publicaties en niet voor datasets, maar de KB wil zich zeker als facilitator blijven opstellen en meedenken in de verdere ontwikkelingen rondom data in relatie tot publicaties om daarmee vanuit de KB een bijdrage te leveren aan de optimale ontwikkeling van de Nederlandse kennisinfrastructuur. Op dit moment vinden gesprekken plaats tussen KB, DANS, 3TU en SURF om dit verder uit te werken.

4.4 Organisatie

De SURFshare community kent een vrij hoge organisatiegraad, met diverse vaste groepen (BIK, WRM, WISH) en diverse werkgroepen. Voor het thema VPs ligt de verantwoordelijkheid voor beleid, organisatie en techniek bijna overal bij de universiteitsbibliotheken. Er is een kleine, maar sterke voorhoede die de nieuwste ontwikkelingen trekt (bijv. via de VP tenders) en ook internationaal actief participeert in VP ontwikkelingen. Deze groep is binnen de SURFshare community erg klein: je komt steeds dezelfde namen tegen.

Het is al langer bekend dat sommige leden van de SURFshare community sneller geneigd zijn om nieuwe zaken te implementeren dan andere. Zo kost het invoeren van de hierboven besproken

¹ Volgens opgave Jeroen Rombouts van 3TU, november 2009

versie 3.0 tijd en energie die men wellicht liever aan andere dingen had willen besteden. In de WRM (Werkgroep Repository Managers) laat de worsteling tussen innovatie en productie zich het sterkst voelen.

Uit gesprekken met leden van de SURFshare community blijkt dat er een toenemende behoefte aan afbakening is. Het begrip VPs wordt als te breed ervaren en daardoor ontstaat het risico dat de complexiteit te groot wordt. Dus: in de beperking toont zich de meester. Dit kan vorm krijgen door nadere afspraken te maken over het begrip VP, bijvoorbeeld door in eerste instantie het oog vooral te richten op eenvoudige VPs die bestaan uit drie afgebakende en identificeerbare onderdelen: de data, de syntax (hoe lees je de data) en het artikel. SURF heeft voor een eerste afbakening van de technische vraagstukken al een begin gemaakt, maar er is nog aanvullend werk nodig.

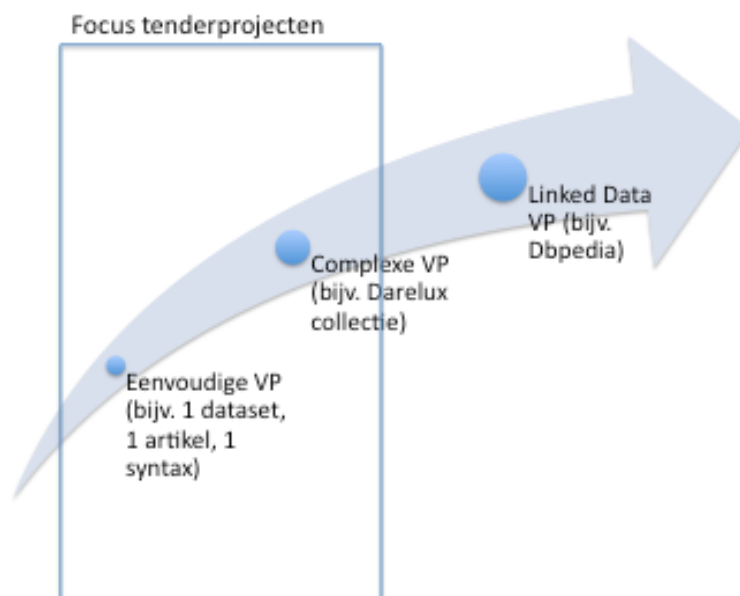
De samenhang tussen de verschillende SURFshare werkpakketten die "iets" doen met het thema VPs is niet helder. In hun opzet zijn de werkpakketten 1 (VPs) en 4 (infrastructuur) bedoeld om elkaar te versterken. De term infrastructuur is binnen deze context niet gericht op het maken van software of technologie, maar veel meer op het maken van afspraken en richtlijnen over toepassing van techniek die er al is. In het verlengde hiervan richt werkpakket 6 zich op de langdurige opslag van onderzoeksdata.

4.5 De tenderprojecten 2009

In het kader van deze rapportage zijn gesprekken gevoerd met de vijf [tenderprojecten](#) die in 2009 zijn uitgevoerd rondom het thema VPs:

- JALC
- Veteran Tapes
- Dataplus
- Escape
- Proefschriften Plus

Hoewel de vijf projecten nog niet geheel zijn afgerond is het wel mogelijk alvast een aantal voorlopige observaties (gebaseerd op gesprekken met de projectleiders van de tenderprojecten) te beschrijven. Voor een verslag van de gesprekken zie bijlage 6.



De projecten bevinden zich ieder afzonderlijk op een onderdeel van een glijdende schaal die loopt van "eenvoudige" VP naar high end semantic web. De nadruk ligt op de eerste twee varianten. Semantic web toepassingen zijn in de tenders nog niet aan de orde gekomen. Dat was en is ook geen onderdeel van het huidige, lopende SURFshare programma.

Het maken van VPs kost (nog) veel tijd en hangt samen met ambitieniveau en vertrekpunt. Twee projecten laten zien dat die tijd via tools (resource map editors) terug gebracht kan worden tot 15 à 30 minuten per VP. Dit is echter in een ideale en afgebakende omgeving. In die verwerkingstijd zijn andere handelingen die vooraf of achteraf aanvullend nodig zijn, niet meegerekend. Denk daarbij aan het vastleggen van een Persistent Identifier (PID) of het identificeren van de auteurs door middel van hun DAI. Maar op het moment dat er sprake is van aanmaak en beheer van vocabulaires en ontologieën gaat de benodigde tijdsinvestering drastisch omhoog. De vraag is wie die tijd er in zou willen steken. De onderzoekers zelf zitten daar over het algemeen niet op te wachten omdat ze hun tijd liever besteden aan hun core business; onderzoek doen.

De manier waarop de relatie tussen artikel en dataset wordt gelegd heeft grote gevolgen. Bijvoorbeeld door het leggen van een link tussen een specifieke paragraaf en (deel van) een dataset, of het linken van (delen van) datasets onderling, of het linken van versies van (delen van) datasets met andere delen van de VP. De complexiteit kan enorm toenemen naarmate de relaties in aantal en diepte toenemen. Het een op een linken tussen artikel en data is een stuk eenvoudiger. Hoe "dieper" de verrijkingen in het artikel gelinkt worden, hoe rijker de functionaliteit voor de gebruiker, maar hoe groter ook de complexiteit van de VP. JALC heeft hier al ervaring mee opgedaan.

In disciplines waar gewerkt wordt met vragenlijsten en statistiek, zoals bijv. sociologie of economie, is de syntax een belangrijk onderdeel van VPs. De syntax vormt dan de sleutel voor correcte interpretatie van de dataset. Zonder deze sleutel is het opslaan van de data wel mogelijk, maar zinloos omdat er geen sprake kan zijn van hergebruik. Zo kan een bestand met ruwe statistische gegevens op een bepaalde manier, omschreven in de syntax (statistische formule), door de onderzoeker zijn bewerkt en daarna geïnterpreteerd.

Producten, workflows, aanpak en dergelijke zijn discipline specifiek en niet zonder meer generaliseerbaar. Dit geldt ook voor metadata.

Kansen en belemmeringen lijken erg op de ervaringen die in het SURF project LOREnet zijn opgedaan. Jane Hunter schrijft over de belemmeringen n.a.v. het Scope project o.a. het volgende:

"...But there are a number of barriers that need to be overcome, before many scientists willingly publish their raw and derived datasets. These include: a lack of simple tools for publishing data with provenance information; lack of motivation for scientists to spend time and effort preparing their data for publication; concern with intellectual property rights; a lack of standards for publishing datasets and provenance; and discipline-specific tools that prohibit cross-disciplinary sharing and exchange. Hence the primary objective of this work is to develop an intuitive, simple, easy-to-use system that enables scientists to quickly author scientific compound objects with built-in provenance and to publish them to a repository with associated metadata and a Creative Commons license – we call this the SCOPE (Scientific Compound Object Publishing and Editing) system. If SCOPE can deliver on these objectives, then the system overcomes some of the current barriers to scientific data publication including: a lack of incentive; lack of tools; difficulty preparing data for publication; difficulty providing an appropriate level of provenance data; concern with intellectual property rights..."

ORE is op zichzelf goed bruikbaar maar er is nog veel pionierswerk, finetuning, en pragmatisme nodig om tot bruikbare resultaten te komen die ook opgeschaald kunnen worden. Net als in de begintijd van OAI-PMH zijn nog veel aanvullende afspraken nodig en die kunnen alleen maar uit de praktische toepassing van ORE te voorschijn komen. Wat dat betreft is de ervaring binnen de SURF

tenders nog mager: per eind 2009 zijn zeven tenderprojecten uitgevoerd. Ieder project ziet een deel van het totale probleemgebied en trekt daar zijn eigen, "lokale" conclusies uit.

Naast een behoefte aan meer praktijkervaring is het nodig om afspraken te maken tussen partijen. Uitgevers en / of bibliotheken (repository managers) moeten regels gaan opstellen voor aanlevering en structurering van VPs. Wederzijdse verantwoordelijkheden moeten vastgelegd worden tussen onderzoeker, universiteitsbibliotheek, uitgever, archief, repository manager enz. Hoe minder de onderzoeker zelf hoeft te doen hoe beter. Daarom moet de workflow en werkverdeling beschreven worden.

De bestaande infrastructuur kan grotendeels generiek gebruikt worden. Tools, data, metadata, workflow en diensten zijn meer discipline specifiek. Een apart probleem vormt het feit dat de aanwezige repositories (nog) niet geschikt zijn voor verwerking van VPs.

In DARE lag de nadruk vooral op disseminatie aan de voorkant. Wat ontbreekt is een centraal loket aan de achterkant waar de onderzoeker heel laagdrempelig in één keer zijn data plus artikel kan aanleveren in ruil voor meerwaarde aan de voorkant. Of zou een CRIS zo'n loket kunnen zijn?

4.6 Andere nationale ontwikkelingen

4.6.1 SURFfoundation

In oktober 2009 organiseerde SURFfoundation de Open Access week. Hoewel de nadruk in deze week niet zozeer lag op het thema VPs zijn er toch belangrijke opmerkingen gemaakt die relevant zijn voor het thema onderzoeksdata (als onderdeel van een VP). Een samenvatting van de belangrijkste opmerkingen staat in bijlage twee. Hieruit blijkt dat er in toenemende mate wordt nagedacht over de manier waarop we in Nederland willen omgaan met onderzoeksdata.

[Onderzoeksdata Forum](#): Onder leiding van SURFfoundation is een Forum opgericht dat samenwerkt aan de opslag en uitwisseling van onderzoeksdata. Het Onderzoeksdata Forum streeft naar verbeterd management van en een betere toegang tot onderzoeksdata voor de wetenschap en maatschappij. De samenwerking in het Onderzoeksdata Forum voorkomt dubbel werk en het Forum kan activiteiten goed op elkaar afstemmen. In deze samenwerking worden de initiatieven binnen diverse organisaties gebundeld: [DANS](#), [3TU](#), [Universiteit van Tilburg](#), de Nationale Coalitie Digitale Duurzaamheid ([NCDD](#)) en SURFfoundation. Concrete resultaten zijn tot op dit moment: Het rapport '[De juridische status van ruwe data; een wegwijzer voor de onderzoekspraktijk](#)' met een [handige verkorte wegwijzer](#). Verder is er het [eindrapport](#) van het door 3TU uitgevoerde project Waardevolle data en diensten

4.6.2 Nederlandse Vereniging voor Beroepsbeoefenaren in de Bibliotheek-, Informatie- en Kennisector (NVB)

In oktober 2009 organiseerde de NVB een workshop over onderzoeksdata. Hoewel er geen nieuwe inzichten zijn bereikt was een van de hoofdzaken die in de discussie naar voren kwam het feit dat het verrijken van publicaties met datasets pas gaat gebeuren als dat een concrete beloning voor de onderzoeker oplevert (met name citaties). Daarnaast was de hoge opkomst (circa 35 personen op de vrijdagmiddag) opvallend. Er is dus kennelijk interesse in het onderwerp: "de rol van bibliotheken bij onderzoeksdata".

4.6.3 Rijksoverheid

Het Ministerie van VWS hanteert een policy voor uitbesteed onderzoek waarbij 10% van de totale subsidie wordt ingehouden tot na het deponeren van de data.

"..De rijksoverheid heeft voor beleidsgericht onderzoek de Model onderzoeksovereenkomst ARVODI-2008 vastgesteld, waarin het deponeren door onderzoekers van onderzoeksgegevens bij

DANS verplicht is gesteld...² [Arvodi](#) staat voor: Algemene Rijksvoorwaarden voor het verstrekken van Opdrachten tot het verrichten van Diensten 2008 (ARVODI-2008)

² <http://www.opendataspeakerscorner.nl/?p=97&lang=nl>

5 Huidige situatie: Internationale ontwikkelingen

Bijlage drie bevat een uitstekend overzicht van diverse internationale ontwikkelingen, opgesteld door Arjan Hogenaar van KNAW. In de samenvatting schrijft hij:

"...Het is zeker voor de wetenschapper verheugend te constateren dat er op zo veel plaatsen op de wereld aan de vernieuwing en verbetering van het wetenschappelijk publicatie- en communicatieproces wordt gewerkt. Na jarenlange stilte wordt er plotseling overal geëxperimenteerd. Toch lijkt het niet onverstandig om op nationaal en zeker op Europees beleidsniveau kritisch te gaan kijken wat er precies gaande is. Het zou een goede zaak zijn als al degenen die zich met die vernieuwing van communicatie bezighielden zich zelf ook zouden houden aan wat ze beogen te propageren: internationaal samenwerken waar dat kan om dubbel werk zo veel mogelijk te voorkomen. De oprichting van de Confederation of Open Access Repositories³ (COAR) in oktober 2009 zal hierin een belangrijke rol kunnen gaan spelen..."

In aanvulling op deze samenvatting kunnen we voor het thema VPs de volgende conclusies trekken.

- Nederland zit in de voorhoede van de ontwikkelingen op het gebied van VPs;
- het thema VPs is ook in het buitenland nog nieuw en wordt nog vrij experimenteel benaderd; echte operationele diensten zijn er nauwelijks;
- DRIVER blijft interessant en belangrijk als voorbeeld en mogelijke bron van tools en software;
- LORE en Scope bieden tools en bouwstenen die we in Nederland zouden kunnen (her)gebruiken;
- ANDS biedt qua projectopzet en projectplannen een mooi voorbeeld voor de verdere plannen die SURFfoundation voor de komende twee jaar gaat maken;
- Semantic web en het Linked Data project zijn veelbelovend maar nog zeker geen main stream ontwikkeling;
- veel wetenschappers proberen op hun eigen wijze nieuwe manieren van wetenschappelijke communicatie te vinden; het beeld wat uit al deze initiatieven van onderzoekers naar voren komt is het beeld van zoeken naar nieuwe manieren van wetenschappelijke communicatie, terwijl men tegelijkertijd de klassieke publicatiepatronen nog niet loslaat. De nieuwe manieren zijn vooral in opkomst in de bèta disciplines waar snelheid een grotere rol speelt dan in de alfa en gamma wetenschappen.

³DRIVER Repository (n.d.). Retrieved 10 September 2009 from <http://www.driver-repository.eu/Confederation-for-European-Digital-Repositories.html>

6 Vraag en aanbod

Bij een analyse van de huidige stand van zaken voor VPs komt vanzelf het onderwerp vraag en aanbod naar boven. In de (inter)nationale bronnen over het onderwerp, in de gesprekken met projectleiders van de tenderprojecten en in de discussies over open acces komen vraag en aanbod impliciet en expliciet aan de orde.

Bijlage twee bevat een aantal citaten die tijdens de Open Access week verschenen op de Open Data Speakers Corner van DANS. De citaten zijn geselecteerd op relevantie voor VPs. Hieruit komt naar voren dat er een (latente) vraag is, er zijn kansen, maar misschien nog wel meer belemmeringen. Opmerkelijk is echter dat uit de citaten angst naar boven komt dat de onderzoeksdata net als commercieel gepubliceerde artikelen achter de 'toll gates' van uitgevers zullen verdwijnen. Impliciet ziet men hier een pleidooi voor open access archieven.

Waar blijkt die (latente) vraag uit? Allereerst uit het feit dat onderzoekers regelmatig aan hun uitgever de vraag stellen op welke wijze zij hun kostbare data kunnen deponeren bij de uitgever. Sommige uitgevers spelen daar op in via een Data Policy. Exacte cijfers over aantallen uitgevers zijn (nog) niet bekend. Het is opmerkelijk en misschien wel zorgwekkend dat onderzoekers zich tot uitgevers richten en niet tot de universiteit waar zij werken. Tegelijkertijd komen er functionele vragen uit het veld (gebruikers / onderzoekers / docenten) die het VP thema sterk raken:

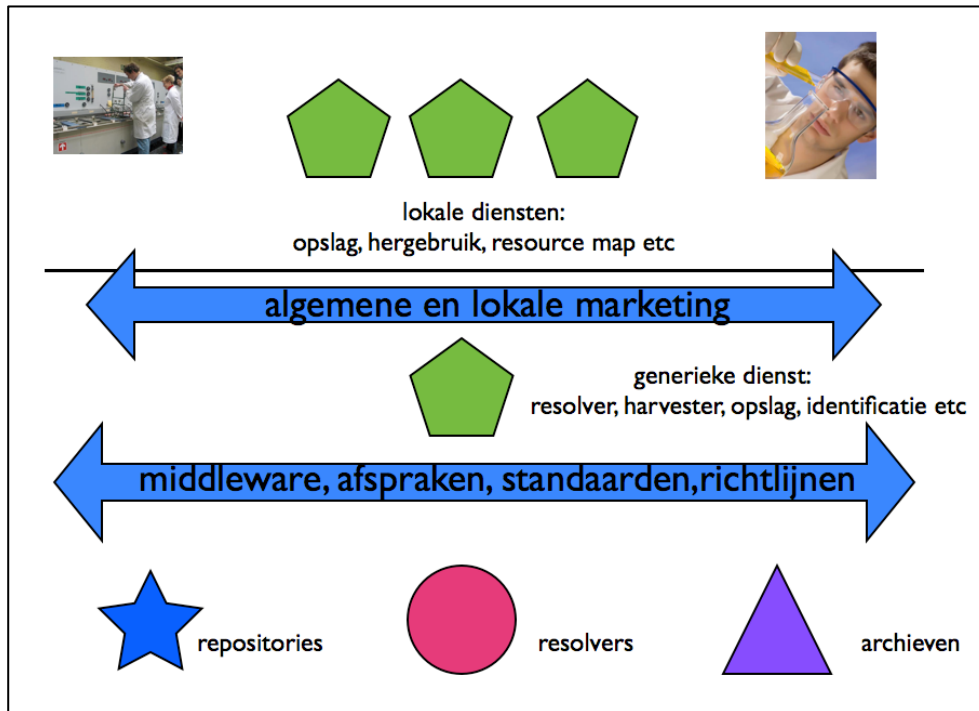
- "Zijn er artikelen geschreven op basis van mijn data? Door wie, wanneer? Maar is mijn data dan correct gebruikt?"
- "Waar kan ik mijn data opslaan? Als de universiteit of de UB dat niet voor me regelt stap ik naar mijn uitgever"
- "Hoe kan ik streaming video of audio bij mijn dissertatie opslaan en gebruiken?"
- "Hoe leg ik een link tussen een tekst- en een audiofragment?"
- "Hoe leg ik de relatie tussen deze presentatie en mijn cursus als geheel vast?"
- "Wij werken met vragenlijsten, aan die vragenlijsten zitten rapportages over validiteit en betrouwbaarheid. Hoe kan iemand die rapporten vinden als die een vragenlijst opvraagt."

Het is belangrijk dat we kijken naar de (latente) vraag. Er is kennelijk behoefte aan ondersteuning bij het onderzoeksproces voor a) het onderling linken van onderdelen b) veilige opslag c) identificatie van bestanden. Er is een vraag en die is niet helder en niet gefocust. Dat is niet zo vreemd want het gaat hier om diffuse verzameling individuen met heel diverse vragen of juist nog zonder ideeën over wat mogelijk is. Helder is in ieder geval dat anderen (SURFshare) voor die helderheid en de focus moeten zorgen, i.c. de vraag articuleren. Tevens moet het aanbod helder gemaakt worden. We kunnen diensten bouwen die inspelen op die vraag. Die diensten moeten we niet "verkopen" met dezelfde terminologie die we intern gebruiken. De term VPs zegt een onderzoeker weinig tot niets. Een dienst die een oplossing biedt voor veilige opslag van zijn kostbare onderzoeksresultaten spreekt waarschijnlijk wel aan.

6.1 Vanuit de onderzoeker bekeken

Het uiteindelijk resulterende model zou kunnen bestaan uit een degelijke infrastructuur (datalaag) waar bovenop een dienst draait die als een soort "groothandel" fungeert op zodanige wijze dat instellingen lokaal de dienst verbijzonderen naar specifieke (lokale) disciplines en groepen. De kern van de dienst is dan generiek, volgens landelijke afspraken en standaarden opgezet, maar de lokale implementatie is specifiek.

Een belangrijk onderdeel van de implementatie van een dergelijk business model is marketing. Een heldere definitie van producten, diensten gerelateerd aan klanten en afnemers is een kritische succesfactor voor de verdere vormgeving van het thema VPs.



Net als bij artikelen en leermiddelen gaat het bij VPs om een cyclisch proces met globaal drie stappen:

1. het archiveren of opslaan
2. het toegankelijk maken
3. hergebruik

De focus lijkt in de huidige aanpak vooral te liggen op de infrastructuur, terwijl het vanuit het perspectief van de wetenschapper gezien belangrijker is om de diensten te regelen. Hergebruik ligt in de meeste gevallen nog ver af van de belevingswereld van de wetenschapper en de weg daar naartoe is bezaaid met belemmeringen, bezwaren en beperkingen. Vooralsnog lijken diensten op het gebied van opslag en toegang het meest kansrijk omdat zij inspelen op de behoefte van onderzoekers aan ondersteuning bij het bewaren, identificeren en zichtbaar maken van hun werk.

7 Conclusies

Als we de huidige situatie in Nederland samenvatten in termen van pluspunten en belemmeringen ontstaat het volgende beeld.

7.1 Pluspunten

Een moderne, solide, state of the art infrastructuur die volop kansen biedt voor verdere ontwikkeling van het thema VPs, met DAI, PID, resolver, repositories, archieven en een solide SURFnet netwerk als voornaamste bouwstenen.

Een relatief goede organisatiestructuur, waarbinnen in goed onderling overleg afspraken worden gemaakt voor toepassing van bestaande standaarden en nieuwe standaarden worden ontwikkeld voor zover nodig.

Een kleine, enthousiaste voorhoede die ondanks zijn kleine omvang nationaal en internationaal impact heeft (zoals o.a. blijkt uit de DRIVER II producten en resultaten). Tools voor het maken van VPs zijn door twee tenderprojecten ontwikkeld (Escape en Dataplus). Volgens de projectleiders bieden deze tools potentieel voor verdere opschaling. Internationaal gezien zijn de meest in het oog springende en zeer waarschijnlijk ook meest bruikbare tools die van ESciDoc, Microsoft Research en Loomp.

Er is redelijk veel materiaal om verder mee te werken: meer dan 10.000 datasets in drie archieven (waarvan 9000 al via Narcis beschikbaar), ca 200.000 open access artikelen, waarvan er circa 130.000 een auteur met een DAI hebben, in totaal bijna 16.000 auteurs met een DAI⁴.

Een redelijk gunstig klimaat waarin de Nederlandse overheid en NWO het deponeren en hergebruik van datasets lijken te willen bevorderen.

7.2 Belemmeringen

VPs zijn conceptueel en in praktische zin een te vaag en te breed begrip. Afbakening en concretisering (discipline gericht) is noodzakelijk.

De IR systemen die nu in gebruik zijn bij Nederlandse instellingen zijn meestal niet geschikt voor verwerking van VPs. Aanpassingen zijn noodzakelijk.

Er is binnen de SURFshare instellingen een vrij grote afstand tussen de kleine voorhoede en de andere instellingen. Het is noodzakelijk om alle instellingen eerst weer op gelijk niveau te brengen alvorens nieuwe stappen te (kunnen) zetten. Er zijn echter voor de SURFshare instellingen geen incentives of sancties om afspraken op te lijnen of op te schonen. Men doet zijn best maar doet het 'erbij'. Daarnaast bestaat er een zekere "angst" dat SURF een te hoog tempo oplegt en de kwaliteit uit het oog verliest.

In de SURFshare community zijn er wel enkele "voormannen" die VPs goed op de kaart kunnen zetten, maar hun aantal is erg klein. De kennis en ervaring moet breder gedeeld worden en de lokale voorzieningen (vooral de DIDL/MODS implementatie) moeten gelijk getrokken worden. Het produceren van VPs kost relatief veel tijd. Wie betaalt die tijd en welk organisatieonderdeel gaat de onderzoeker hiermee helpen? De bibliotheek ligt al zwaar onder vuur door allerlei bezuinigingen.

VPs zijn in meerdere opzichten discipline specifiek. De infrastructuur kan generiek opgezet en benut worden maar op het niveau van tools, workflow en diensten is een specifieke aanpak vereist.

⁴ Bron: opgave KNAW begin november 2009

8 Aanbevelingen

Zonder te vervallen in ongegrond optimisme is het zeer de moeite waard om te onderzoeken in hoeverre de genoemde belemmeringen omgebogen kunnen worden tot kansen en op welke manier dat zou kunnen. Op basis van dit onderzoek en op basis van de uitkomsten van de workshop verrijkte publicaties zijn de volgende zaken relevant:

► **Meet them where they are**

Een mindshift van artikel naar data en het “parkeren” van artikelen als aandachtsgebied zal in de praktijk niet werken. Onderzoekers denken nog steeds vanuit het artikel. We kunnen dus niet anders dan werken vanuit het artikel en de benadering van de onderzoeker rond het thema VPs moet dan ook gebaseerd zijn op het artikel. In de toekomst kan de stap gezet worden naar data driven science, maar nu nog niet.

Door te denken vanuit de behoefte aan ondersteuning bij de onderzoeker kunnen de resultaten van het SURFshare programma als concrete dienst – misschien eerst nog in bèta versie – aangeboden worden op een manier die qua terminologie en inhoud aansluit bij die behoefte. Door het maken van use cases en het ontwikkelen van views op de diensten voor afzonderlijke disciplines kan de meerwaarde van de VP/collaboratory diensten zichtbaar gemaakt worden.

► **Keep it simple**

Uit de ervaringen met de tenderprojecten en opmerkingen uit het veld blijkt dat we niet alles tegelijk kunnen doen. Het begrip VPs is erg breed en daardoor dreigt een te grote complexiteit. Het lijkt verstandig om vanaf nu eerst de aandacht te richten op “eenvoudige VPs” (artikel, syntax, dataset, enkelvoudige relatie). Dit geeft ook meer focus in de aanpak.

► **Er zijn belangrijke infrastructurele zaken die met voorrang aangepakt moeten worden**

Op basis van de “meet them where they are” opmerking mag verwacht worden dat het de onderzoeker zal aanspreken om eerst een oplossing te bieden voor veilige opslag van zijn data, voor permanente identificatie (persistent identifier) van zijn artikel en zijn data in onderlinge samenhang, maar ook in relatie tot de onderzoeker zelf (via de DAI).

1. Voorzieningen voor opslag van data, als onderdeel van de workflow van een onderzoeker, moeten met prioriteit worden onderzocht en als dienst aan de onderzoeker worden aangeboden.
2. Het toevoegen van identifiers (PID, DAI) aan onderdelen van een VP is nog geen standaard procedure en moet daarom eerst als eerste activiteit aan de workflow worden toegevoegd na opslag van die onderdelen.
3. Het maken van landelijke afspraken of richtlijnen over opslag, verwerking, transport enz. van VPs op zodanige wijze dat dataproviders en serviceproviders samen tot werkende processen kunnen komen.

► **Bestaande IR systemen geschikt maken voor VPs**

Omdat de meeste repositories (nog) niet geschikt zijn voor de verwerking van VPs zullen instellingen lokaal maatregelen moeten nemen. De eerste stap bestaat uit het op gelijk niveau brengen van de afspraken. Begin 2010 zullen alle instellingen versie 3.0 in gebruik hebben genomen. Daarmee is transport van VPs van IR naar bijv. Narcis mogelijk. De tweede stap bestaat uit het geschikt maken van het IR systeem voor invoer en opslag van VPs. De lokale repository manager kan daarvoor kiezen uit een aantal opties:

- zelf een workaround aanbrenge;
- een oplossing die als apart platform naast het IR systeem voor VPs wordt gebruikt, bijv. eSciDoc;
- DSpace gebruikers kunnen wachten op de aangekondigde compound object versie van DSpace;
- overstappen van het bestaande IR systeem naar Fedora.

De eerste optie is bijv. in Utrecht in een van de tenderprojecten gebruikt. Daaruit blijkt dat de workaround niet door anderen gekopieerd kan worden. Een workaround is ook niet toekomstvast. De tweede optie lijkt de eenvoudigste maar heeft als nadeel het feit dat het geen geïntegreerde

oplossing is. Voor DSpace gebruikers lijkt geduld het enige wat nodig is, maar de vraag is hoe lang de nieuwe versie op zich zal laten wachten. De vierde optie is weliswaar ingrijpend maar het is bekend dat sommige instellingen een migratie naar Fedora al enige tijd overwegen. Het is denkbaar dat instellingen die zo'n overstap overwegen zich verenigen zodat in samenwerking met SURF toegewerkt kan worden naar een generiek platform dat lokaal gebruikt kan worden voor zowel artikelen als VPs op basis van de actuele standaarden.

Binnen SURF bestaan vergevorderde plannen om het geschikt maken van IR systemen voor VPs te ondersteunen. Daartoe wordt begin 2010 een activiteit gestart onder de naam RING, Repository Infrastructure Next generation.

➤ **Sluit zo vroeg mogelijk aan bij de workflow van een onderzoeker**

Focus op een beperkt aantal generieke onderzoeksmethodieken is, naast focus op kansrijke disciplines zoals rechten, sociologie, archeologie, economie, zeer waarschijnlijk een kansrijke aanpak. Het is belangrijk om in een zo vroeg mogelijk stadium in het onderzoeksproces te starten met data capture om later tot een VP te kunnen komen. We moeten aansluiten bij de methodieken en workflows van de onderzoeker en niets opleggen. Hier ligt een duidelijk raakvlak met het thema Collaboratories aangezien deze virtuele omgevingen met bijbehorende tools bij uitstek de plek zijn waar onderzoekers aan (verrijkte) publicaties kunnen werken

➤ **Kwaliteit is belangrijker dan kwantiteit**

Het is noodzakelijk om voor de komende twee jaar toe te werken naar concrete, zichtbare resultaten. Een aansprekend aantal helpt altijd maar het realiseren van x verrijkte publicaties per y moet niet het leidende principe zijn.

➤ **Beleidsontwikkeling door universiteitsbibliotheken nodig voor dienstverlening aan onderzoeker**

De plannen die SURF gaat ontwikkelen moeten kunnen rekenen op draagvlak en dat moet vooral gevonden worden in de universiteitsbibliotheken. Immers: in bijna alle gevallen is het werk dat samenhangt met VPs belegd bij de UB en niet bij de onderzoekers zelf. Deze constatering leidt tot een andere, meer beleidsmatige observatie:

Repositories zijn een vast onderdeel geworden van de UB dienstverlening. Dienstverlening voor VPs is in ontwikkeling en daarom is tijdelijke ondersteuning nodig en gerechtvaardigd. Ondersteuning (vooral bij het geschikt maken van bestaande IR systemen voor VPs) moet daarom onderdeel zijn van de plannen voor de komende twee jaar. Daarnaast moet de UB zelf beleid gaan ontwikkelen voor de periode na 2011, om op eigen kracht dienstverlening te kunnen garanderen en continueren. VPs zijn geen doel op zich maar een middel om onderzoeksresultaten van de eigen instelling meer zichtbaar te maken en dat dient het hogere doel van kennisdeling en vergroten van de zichtbaarheid van de universiteit of HBO instelling als geheel.

Tijdens de workshop bleek duidelijk dat de meningen en verwachtingen over het thema VPs sterk uiteen lopen. Het onderwerp is voor velen nog erg nieuw en dat bleek ook tijdens de discussies in de workshop. Maar ook onder degenen die al langer met het onderwerp bezig zijn lopen de meningen uiteen. Zoveel mensen, zoveel meningen, waarbij men toch met een schuin oog naar SURF kijkt.

➤ **Leiderschap vanuit SURF nodig**

De leden van SURF zijn SURF, het kan niet vaak genoeg benadrukt worden. Toch is er vooral op uitvoerend niveau behoefte aan richting, inspiratie en leiderschap en men kijkt dan toch naar de mensen van SURF. Die zitten immers niet zo vast in de dagelijkse beslommingen van het universitaire bedrijf en kunnen met een helikopterblik de richting aangeven. Sommige deelnemers aan de workshop gaven expliciet aan dat het thema VPs beter zal gedijen als er meer focus op komt.

Het is belangrijk dat SURF de hoofddoelen van het SURFshare programma, met de vertaling van deze doelen naar de werkpakketten, blijft communiceren. Er is naast afbakening van het thema VPs ook behoefte aan een heldere, herhaalde, consequente boodschap die kort en krachtig laat

zien waar het naar toe gaat. VPs zijn geen doel op zich maar een middel om onderzoeksresultaten meer zichtbaar te maken en meer naar buiten te brengen.

Een belangrijke vraag is: wat gebeurt er als het thema VPs niet tot voldoende resultaat leidt per einde 2011? Voor de plannen 2010-2011 is de vraag wel zeer relevant. Maar wat is "voldoende resultaat"? Hoe belangrijk zijn VPs voor het succes van het SURFshare programma? Willen de SURF leden een vervolg op het SURFshare programma en zo ja: wat moet er dan nu al gebeuren met VPs om dat vervolg succesvol te laten zijn? Hoe moeilijk de beantwoording ook moge zijn, afwachten kan niet.

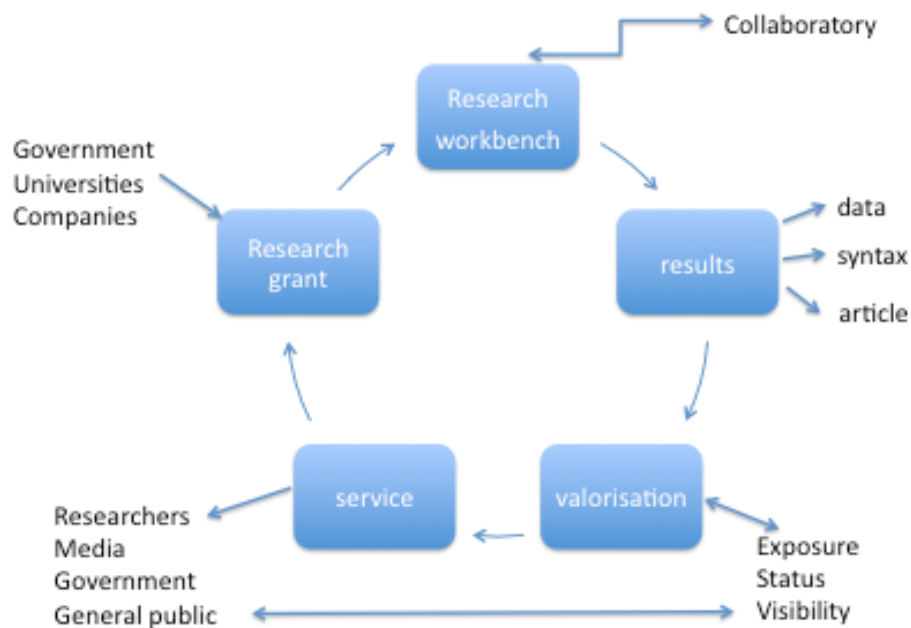
► **Een zichtbare, inspirerende dienst is een voorwaarde voor een vervolg op SURFshare**

Eind 2011 komt het lopende SURFshare programma tot een afronding. We weten nu al zeker dat het werken aan de innovatie van wetenschappelijke communicatie daarmee niet is afgerond. Een vervolg is nodig. Innovatie is ook geen doel op zich maar een middel om de onderzoeker optimaal te ondersteunen. Zichtbare dienstverlening die inspeelt op de vragen van de onderzoeker is daarom onontbeerlijk. Die dienstverlening kent twee complementaire onderdelen:

1. De landelijke infrastructuur, (verder) te ontwikkelen door de SURFshare instellingen (UB en HBO mediatheek) op basis van gemeenschappelijke afspraken, richtlijnen en aanpak onder leiding van en ondersteund door SURFfoundation. De resultaten van deze aanpak komen eind 2011 samen in een portal of website die de dan beschikbare diensten voor onderzoekers als operationele dienst laat zien in de vorm van een substantieel aantal VPs, operationele diensten voor opslag van data, operationele diensten voor identificatie van data, artikelen en personen, uitleg over en voorbeelden van werkende tools en workflows binnen een aantal disciplines, en links naar de lokaal per instelling aangeboden diensten.
2. De lokale infrastructuur, (verder) te ontwikkelen door de SURFshare instellingen (UB en HBO mediatheek), op basis van de landelijke afspraken, maar toegespitst op de lokale behoefte van vakgroepen, faculteiten, onderzoeksscholen, promovendi. De resultaten van deze aanpak zijn eind 2011 zichtbaar in de lokale dienstenportfolio die zou kunnen bestaan uit bijvoorbeeld een loket waar de onderzoeker terecht kan voor opslag van data, toekennen van identifiers, hulp bij het linken van data aan artikelen (met behulp van de tools die door de tenderprojecten zijn ontwikkeld of tools die ontleend worden uit buitenlandse projecten als eSciDoc), enzovoorts. Lokaal zullen hier verschillen in ontstaan maar dat is geen nadeel, eerder een voordeel.

Bij de ontwikkeling van deze infrastructuur en daarop gebaseerde landelijke en lokale diensten moeten de hierboven genoemde observaties worden meegenomen.

Een tweejarig project onder de (voorlopige) titel "VPs 2011" zou invulling kunnen geven aan de realisatie van een aansprekende, zichtbare dienst die de voordelen van opslag, identificatie en zichtbaarheid van onderzoeksdata via VPs concreet aantoont. De aanpak kenmerkt zich daarbij vooral door het werken met bestaande middelen en organisatie. Het gaat primair om het benutten van aanwezig potentieel in de aanwezige infrastructuur en niet om het ontwikkelen van nieuwe technologie. Kwaliteit is daarbij belangrijker dan kwantiteit.



Naast collaboratories zijn VPs het enige onderdeel van het lopende SURFshare programma waarmee draagvlak kan ontstaan voor een mogelijk vervolg op SURFshare. Beleidsmatige (her)ijking van het belang van de afzonderlijke componenten van het SURFshare programma en hun bijdrage aan het succes van het totale programma moet nu plaatsvinden. Op basis daarvan moeten resources naar rato van hun belang worden toegekend en aandacht krijgen in de komende twee jaar. Focus op VPs en collaboratories als onderdeel van en als middel tot ontwikkeling van diensten voor onderzoekers. De koers voor de komende twee jaar moet zo snel mogelijk worden vastgesteld en daarna gecommuniceerd worden naar de SURFshare community.

Bijlage 1 - Bronnen

Nationale ontwikkelingen

Veteran Tapes

Contactpersonen: Paula Witkamp - paula.witkamp@bureau.knaw.nl
Rene van Horik - rene.van.horik@dans.knaw.nl

Proefschriften Plus

Contactpersoon: Martin Slabbertje - m.slabbertje@uu.nl

JALC

Contactpersoon: Henriette Reerink - h.a.reerink@uva.nl

Dataplus

Contactpersonen: Ellen Verbakel - E.Verbakel@uvt.nl
Ruud Luijkx - R.Luijkx@uvt.nl

ESCAPE

Contactpersonen: Esther Hoorn - e.hoorn@rug.nl
Henk Ellerman - h.h.ellermann@ub.rug.nl

Voor de bovenstaande tenderprojecten is een [wiki](#) aangemaakt.

Het artikel Enhanced Publications: Defining the next steps / Magchiel Bijsterbosch

Internationale ontwikkelingen

Australië

de [Aus-e-lit tool](#) en het [ANDS](#) project.

DRIVER

[Driver rapporten](#) rondom verrijkte publicaties zijn opgeleverd in 2008 en gereviewed in 2009.

Knowledge Exchange

Het Knowledge Exchange "Enhanced E-theses project" geschreven door Martin Slabbertje

Bijlage 2 - Open data

Samenvatting van discussies op [Open Data Speakers Corner](#) Van DANS.

„Deze data moeten echter wel op een zodanige manier worden aangeleverd dat hergebruik mogelijk is en daar schort het momenteel aan. Met het door DANS gerealiseerde EASY-instrument voor Open Access is een belangrijke stap gezet. Maar totdat het op de juiste wijze deponeren van onderzoeksdata een wezenlijk onderdeel is van het universitaire opleidingsprogramma, is het zeker ook de taak van universiteitsbibliotheken om onderzoekers te ondersteunen bij het uploaden van hun data en artikelen in de betreffende repositories...”

Makkelijke toegang tot de datasets is echter niet genoeg. Nog steeds nodig zijn bewezen strategieën voor het creëren van zinvolle koppelingen tussen data. Hoe meer connecties en relaties er kunnen worden aangebracht tussen openbare datasets, hoe steviger de funderingen voor wetenschappers om lagen aan te brengen, uit te breiden en erop voort te bouwen. Bovendien kan het hogere aantal koppelingen helpen om een licht te doen schijnen op innovatieve toepassingen die mogelijk niet zichtbaar zijn als de data afzonderlijk worden onderzocht. Als gegevens eenmaal kunnen worden vermengd met andere onderzoeksinformatie, zal dit tot veel wetenschappelijke innovaties leiden. Een andere uitdaging van open data is de aarzeling bij wetenschappers om hun met pijn en moeite verkregen data te delen met concurrerende onderzoekers. Er begint een dualiteit zichtbaar te worden in de wetenschappelijke gemeenschap van meer samenwerking in een omgeving waarin de concurrentiedruk steeds hoger wordt. Velen worstelen om de juiste balans te vinden. Kunnen beloningen als academische “credits” voor wetenschappers worden gemaakt om het delen van data aan te moedigen? Rafael Sidi, vice president of product development for Elsevier’s ScienceDirect

Openheid van data zal er op termijn zeker van komen. Nu al is het bij sommige tijdschriften in de levenswetenschappen verplicht om de data openbaar te maken. In andere domeinen staat men echter nog niet te springen om volledige openheid en wordt het bezit van de data gezien als een onderdeel van de concurrentiestrijd. Overigens, daar waar sprake is van onderling vertrouwen op basis van samenwerking worden data gedeeld. Het stimuleren van samenwerking is om die reden dus al belangrijk. Belangrijk is verder dat openheid nog niet hetzelfde is als bruikbaarheid. Dit is bij een instituut als DANS (KNAW-NWO instituut), waar ze bij voortduring aan dataverrijking doen, maar al te goed bekend. Lees ik nu door de regels heen de gedachte dat uitgever ook datadepots kunnen opzetten (zoals bv Nature doet)? Zou dat wat zijn voor Elsevier? Dit brengt gemengde gedachten teweeg. Data behoren openbaar te zijn, maar wil je ze bij een commerciële uitgever opslaan of toch liever bij DANS? Theo Mulder, Directeur Onderzoek en Instituten, KNAW. Een publicatie hangt in toenemende mate van de onderliggende data af en kan niet goed genoeg worden beoordeeld of vol vertrouwen als basis worden gebruikt om verder mee te werken zonder toegang te hebben tot die data. Bovendien zijn in sommige gevallen de gegevens zelf net zo zeer een eersteklas onderzoeksresultaat als de publicatie. Maar voor wie gegevens schept en beheert, hebben nog niet dezelfde soort beloningen als voor degenen die artikelen en papers voor congressen schrijven. Evenmin hebben we eenzelfde kwaliteit infrastructuur voor gegevens (met noemenswaardige uitzonderingen, zoals DANS) als we voor artikelen hebben, met zaken zoals repositories op het instituut en innovatieve interfaces zoals NARCIS. Andrew Treloar. Onderdirecteur, Australian National Data Service.

Daarnaast behoort dit materiaal ook toegankelijk te zijn voor de computer als ‘lezer’, als onderzoeksinstrument. Het digitaal materiaal dient onttrokken te worden aan de digitale silo’s waarin het opgesloten ligt. Het op grote schaal kunnen indexeren, doorzoeken, koppelen, manipuleren en masseren van digitale informatie maakt nieuwe vormen van wetenschappelijk onderzoek mogelijk. Hierbij kunnen technieken zoals text- en datamining ingezet worden. [...]Maar open access was toch niet alleen hiervoor bedoeld? Open access als publicatie- of businessmodel had/heeft ook de ambitie om wetenschappelijke informatie betaalbaar te houden. Kurt De Belder, Bibliothecaris van de Universiteit, Directeur Universitaire Bibliotheken Leiden & Leiden University Press

Hergebruik, ook door commerciële partijen, is van groot belang om de vindbaarheid van bronnen te optimaliseren. De licentiëring van open access artikelen dient op dit punt niet te beperkend te zijn. Ik verwijs hierbij graag naar het initiatief tot de gratis site <http://www.liigl.nl>. **Dr. L. (Laurens) Mommers**, Faculteit Rechtsgeleerdheid, Leiden

Het door NWO gefinancierde MESS project met als kernonderdeel een representatief panel van Nederlandse huishoudens (het LISS panel) stelt alle verzamelde data zo snel mogelijk beschikbaar aan iedere onderzoeker die als doel heeft data te gebruiken voor wetenschappelijk, beleids- en maatschappelijk relevant onderzoek (<http://www.lissdata.nl>). Hopelijk volgen meerdere onderzoeks- en overheidsinstellingen dit initiatief. Ik juich het van harte toe! Marcel Das, CentERdata, Universiteit van Tilburg

Datasets (en zeker de registerbestanden van de overheid) zijn alleen maar interessant als deze niet alleen beschikbaar worden gesteld, maar vervolgens ook goed onderhouden worden. Actualiteit, betrouwbaarheid en mogelijkheid tot terugmelding op data zijn essentiële onderdelen van datasets: Een dataset moet 'levend' zijn. Leven kost geld. Dat betekent dat er lange termijn budget gereserveerd moet worden, of inkomsten gegenereerd moeten worden met de data zelf. In de praktijk blijkt dat afspraken hierover het moeilijkste onderdeel zijn van beschikbaarstelling van data. **Kees de Zeeuw, manager product- en procesinnovatie GEO, Kadaster**

Databestanden moeten er wel zijn. Iemand moet ze aanleggen. Epidemiologische gegevens over heel wat landen en aandoeningen ontbreken of worden niet systematisch bijgehouden. We kennen zelfs zogenaamde "verwaarloosde ziekten" – leptospirose, chagas of leishmaniasis om er een paar te noemen. Deze ziekten worden niet verwaarloosd omdat ze niet belangrijk zijn – het tegendeel is het geval. Maar we weten niet hoeveel miljoenen(!) mensen er aan lijden. Op basis van goede epidemiologische data en hun verspreiding zou het niet moeilijk zijn om te besluiten, dat we onmiddellijk moeten investeren in goede therapieën! dr. Jan Donner, directeur Koninklijk Instituut voor de Tropen

Bij het CBS is het al een aantal jaren mogelijk om gebruik te maken van databestanden op individueel niveau (microbestanden). Sinds enkele jaren kan dat ook via een remote access verbinding, waardoor onderzoekers vanaf de eigen werkplek analyses kunnen uitvoeren. Hier is veel belangstelling voor, ook al omdat het mogelijk is zelf databestanden in te brengen en die op individueel niveau te koppelen met CBS bestanden. Maar het wordt nog veel aantrekkelijker als je bedenkt dat dit ook met databestanden van anderen mogelijk is. Om uitwisseling te bevorderen zou je er aan kunnen denken een stelregel in te voeren dat bestanden van anderen te gebruiken zijn als je zelf iets inbrengt. Als je daarnaast nog bedenkt dat uitvoering van het DISS-project (Data Infrastructure for the Social Sciences) betekent dat deze voorziening voor een belangrijke groep onderzoekers vrijwel gratis te benutten zou zijn, dan lonken er wel hele mooie perspectieven. Frans Hoeve, Centrum voor Beleidsstatistiek, CBS

Bijlage 3 - Overzicht internationale ontwikkelingen

Het onderstaande overzicht is grotendeels gebaseerd op het overzicht van internationale ontwikkelingen dat is gemaakt door Arjan Hogenaar van de KNAW in het kader van de SURF publicatie over het thema Verrijkte Publicaties.

DRIVER

De initiatieven die in Nederland zijn ontplooid op het terrein van verrijkte publicaties staan niet op zich zelf. Deze zijn voor een belangrijk deel gerelateerd aan wat in de DRIVER community aan vernieuwende ideeën is uitgedacht. In dit verband kunnen de DRIVER studies⁵ kunnen genoemd, waarbij met name Muriel Foulonneau en Francis André in hun rapport met de titel 'Investigative Study of Standards for Digital Repositories and Related Services'⁶ al een eerste heldere beschrijving van het toegang bieden tot verrijkte publicaties geven. In DRIVER-II stond het Nederlandse aandeel in het projectwerk bijna geheel in het teken van VPs. Dat werk is in 2008 afgerond. In de loop van 2009 is gewerkt aan het uitbreiden van de binnen DRIVER ontwikkelde Dnet software met een *compound object service* die geschikt is voor verwerking van VPs. Oplevering van deze service zal in de komende maanden plaatsvinden. Hoewel de DRIVER projecten niet onder die naam worden voortgezet, ziet het ernaar uit dat in vervolprojecten aandacht aan VPs geschonken blijft worden. Het is daarom belangrijk dat SURFshare contact houdt met de partners in deze projecten.

eSciDoc

In Duitsland ontstond onder de onderzoekers van de Max Planck Gesellschaft⁷ (MPG) enige jaren geleden de behoefte aan nieuwe wijzen van wetenschappelijke communicatie. Dankzij de omvang van het MPG kunnen vernieuwende programma's in eigen huis worden gestart. Het bekendste programma dat relevant is voor de wereld van Verrijkte Publicaties is eScienceDoc⁸. Dit project in een samenwerkingsverband van MPG en FIZ-Karlsruhe. eSciDoc richt zich op de behoefte van onderzoekers om direct toegang te hebben tot allerlei soorten informatieobjecten (publicaties; afbeeldingen; datasets; annotaties; commentaren).

De verschillende informatietypen kan de onderzoeker samenbrengen in scientific knowledge space. De - permanent toegankelijke - onderdelen van het samengestelde object worden aan elkaar gerelateerd⁹. EsCiDoc is opgetuigd met alle moderne snuffjes die collaborative work en de productie van EP's mogelijk maken.

DARIAH

Digital Research Infrastructure for the Arts and Humanities¹⁰ (DARIAH) is een omvangrijk Europees project dat deels geïnspireerd is door astronomisch onderzoek. Vooral binnen de zogenaamde 'digital humanities' (ook wel 'computational humanities') speelt een soortgelijke problematiek als binnen de bètavakken. Het gaat om de toegang tot gedistribueerde informatie met een veelvoud aan beschikbare informatietypen en om het uitgeven van publicaties waarin de kerntekst gerelateerd is aan de data waarop die tekst gebaseerd is.

⁵ Zie: DRIVER Repository (n.d.). Retrieved 10 September 2009 from <http://validator.driver.research-infrastructures.eu/validatorWeb/>

⁶ Zie: Foulonneau, M. & André, F. (2007). Investigative study of standards for Digital Repositories and related services (2007). Driver Study 3. Amsterdam UP. ISBN 9789053564127. 112 p. Retrieved 10 September 2009 from <http://dare.uva.nl/aup/nl/record/260226/>

⁷ Zie: Max-Planck-Gesellschaft (n.d.). Retrieved 10 September 2009 from <http://www.mpg.de/>

⁸ Zie: eSciDoc Project (n.d.). Retrieved 10 September 2009 from <https://www.escidoc.org/>

⁹ Zie: Dreyer, M., Bulatovic, N., Tschida, U. & Razum, M. (2007). eSciDoc: a Scholarly Information and Communication Platform for the Max Planck Society (2007). Retrieved 10 September 2009 from <https://www.escidoc.org/media/docs/ges-eSciDoc-article.pdf>

¹⁰ Zie: DARIAH (n.d.). Retrieved 10 September 2009 from <http://www.dariah.eu/>

In DARIAH wordt gewerkt aan een infrastructuur die de computational humanities (voor een mooie inleiding op dit onderwerp zie de 'Companion to Digital Humanities'¹¹) een flinke impuls gaat geven.

Dart and Archer at Monash University

Op de Monash University lijken de ontwikkelingen op die bij het Max Planck Gesellschaft. De onderzoekers hebben enkele jaren geleden aangegeven dat een nieuwe infrastructuur voor eScience wenselijk was. Nadat er repositories waren ingericht bleek er namelijk behoefte aan middelen en technieken om alle aan het onderzoeksproces gerelateerde informatie vast te leggen (experimenten; datasets, publicaties en leermiddelen). Op grond van deze behoefte is eerst het project DART¹² (Dataset Acquisition, Accessibility, and Annotation e-Research Technologies) opgezet. Andrew Treloar heeft een beknopte omschrijving¹³ van het project opgesteld. DART heeft de bouwstenen geleverd waarop inmiddels is voortgeborduurd in het project ARCHER¹⁴ (Australian ResearCH Enabling enviRonment). ARCHER is nog iets verder gegaan dan DART omdat het aangepast kan worden aan de behoeften van de afzonderlijke disciplines. Binnen de sociale wetenschappen kan ARCHER dus anders worden ingericht dan in de natuurwetenschappen, zonder dat de basis infrastructuur wordt aangepast. ARCHER zelf biedt een gedetailleerd overzicht¹⁵.

SCOPE

De Australische Jane Hunter heeft met haar artikel over Scientific Publication Packages¹⁶ de SURFshare Community geïnspireerd om te gaan werken aan tools om Verrijkte Publicaties te produceren. Een praktisch vervolg van haar artikel is het project SCOPE¹⁷: Scientific Compound Object Publishing and Editing. De doelstelling van SCOPE lijkt op die van ESCAPE en DATAPLUS in de SURF tender projecten: het bieden van authoring tools aan onderzoekers zodat deze zelf verrijkte publicaties kunnen samenstellen. Gebruikers kunnen via een embedded web browser zelf objecten op internet vinden en toevoegen aan hun verrijkte publicatie. Hierop voortbouwend is Literature Object Re-use and Exchange (LORE)¹⁸ ontwikkeld dat ongeveer dezelfde functionaliteit biedt, maar nu vooral gericht op de literatuurwetenschap. LORE maakt gebruik van een ontologie die is afgeleid van het door IFLA opgestelde FRBR (Functional Requirements for Bibliographic Records). Een recent voorbeeld van een binnen LORE ontwikkelde VP viewer en editor is te zien op <http://blip.tv/file/2132990/>

myExperiment

Veelbelovend in Engeland is de myExperiment¹⁹ omgeving. myExperiment wordt omschreven als een 'Social Virtual Research Environment'. Aanleiding om myExperiment op te zetten is het feit dat het voor onderzoekers steeds lastiger wordt om in de veelheid aan literatuur de juiste artikelen te vinden. Binnen myExperiment wordt het auteurs gemakkelijk gemaakt om op moderne wijze het gehele wetenschappelijke productieproces weer te geven: workflow, data en publicaties inclusief hun onderlinge relaties.

¹¹ Zie: Schreibman, S., Siemens, R. & Unsworth, J. (eds.). (2004). A Companion to Digital Humanities. Oxford, Blackwell. ISBN: 9781405103213. Retrieved 10 September 2009 from <http://www.digitalhumanities.org/companion/>

¹² Zie: DART (n.d.). Retrieved 10 September 2009 from <http://dart.edu.au/>

¹³ Zie: Treloar, A. (2006). The DART Project: building the new collaborative e-research infrastructure. The 12th Australasian World Wide Web Conference, 1-5 July, 2006. Retrieved 10 September 2009 from <http://ausweb.scu.edu.au/aw06/papers/refereed/treloar/paper.html>

¹⁴ Zie: ARCHER (n.d.). Retrieved 10 September 2009 from <http://archer.edu.au/>

¹⁵ Zie: Bennett, S. (2008). Archer 1.0 Overview. Retrieved 10 September 2009 from <http://archer.edu.au/products/archer-overview.pdf>

¹⁶ Zie: Hunter J. (2006). Scientific Publication Packages – a Selective Approach to the Communication and Archival of Scientific Output. *International Journal of Digital Curation*, 1 (1). Retrieved 10 September, 2009 from <http://www.ijdc.net/index.php/ijdc/article/view/8/4>

¹⁷ Zie: Cheung, K., Hunter, J., Lashtabeg, A. & Drennan, J. (2008). SCOPE: A Scientific Compound Object Publishing and Editing System. *International Journal of Digital Curation*, 3 (2). Retrieved 10 September, 2009 from <http://www.ijdc.net/index.php/ijdc/article/view/84/55>

¹⁸ Zie: Gerver, A. & Hunter, J. (2009). LORE: A Compound Object Authoring and Publishing Tool for Literary Scholars based on the FRBR. University of Queensland, Australia. Retrieved 10 September, 2009 from www.itee.uq.edu.au/~eresearch/papers/2009/OR2009.pdf

¹⁹ Zie: myExperiment (n.d.). Retrieved 10 September, 2009 from <http://www.myexperiment.org/>

Net als bij SCOPE krijgt de onderzoeker de mogelijkheid zelf te bepalen welke objecten hij van belang acht voor een bepaalde verrijkte publicatie. Het accent ligt op de uitwisseling en het becommentariëren van informatie-objecten²⁰, waarbij de workflow een sleutelpositie inneemt.

ICE-TheOREM - End to End Semantically Aware eResearch Infrastructure for Theses

Terwijl in Nederland hard wordt gebouwd aan de Enhanced Theses door de Universiteit Utrecht, wordt aan de andere kant van de wereld in Australië gewerkt aan ICE-TheOREM. In ICE-TheOREM worden de functionaliteiten van Integrated Content Environment²¹ en ORE in één systeem aan elkaar gekoppeld. Deze unieke combinatie maakt het onderzoekers mogelijk om zonder al te specifieke technische kennis verrijkte publicaties samen te stellen die tegelijkertijd desgewenst als PDF kunnen worden gepubliceerd. Peter Sefton en Jim Downing hebben tijdens de Open Repositories conferentie in 2009 ICE-TheOREM verder toegelicht²².

ECrystals

Een klassieker in de wereld van verrijkte publicaties is het eCrystals repository²³ project van Peter Murray Rust. eCrystals is een van de belangrijkste voorbeelden geweest bij de ontwikkeling van de DANS/KNAW-OI Demonstrator. Wat maakt eCrystals zo onderscheidend? eCrystals is opgezet lang voordat Herbert van de Sompel met de OAI-ORE concepten kwam en is ontstaan binnen de onderzoekswereld zelf. Kristallografen kregen behoefte aan het relateren van informatie aan kristallogrammen. Zo wordt in eCrystals niet alleen de visualisatie van de ruwe data getoond, maar ook de ruwe data zelf alsmede diverse bibliografische informatietypen. In feite was en is eCrystals een verzameling verrijkte publicaties avant la lettre. Het heeft als voorbeeld gediend voor latere VP-projecten.

Liquid Publications

Tot slot dient in dit korte overzicht het Liquid Publications²⁴ project niet onvermeld te blijven. In het omvangrijke project werken wetenschappelijke instellingen samen met de uitgever Springer. Het idee is het opzetten van een geheel nieuwe infrastructuur die ruimte biedt aan een nieuwe wijze van wetenschappelijk publiceren. Hierbij gaat het niet alleen om het aan elkaar knopen van diverse informatietypen. Onderzoekers kunnen elkaars publicaties becommentariëren, kunnen een samenwerkingsverband aangaan op basis van dit nieuwe publicatietype en kunnen nieuwe informatie aan de publicatie toevoegen²⁵.

Linked Data

[Linked Data](#) is een vrij actieve community met Tim Berners-Lee als meest zichtbare en inspirerende voorman. Meest concrete en zichtbare resultaat is [DBpedia](#). Dit voorbeeld laat duidelijk de voordelen van de Linked Data principes zien voor de gebruiker. Nadeel is ook zichtbaar: DBpedia wordt gemaakt o.b.v. een dump van de "platte" Wikipedia. Dat kost tijd en geld. Welke service providers gaan in de toekomst deze services leveren en waarom? Wat is het business model? Zie hiervoor ook het [artikel](#) over Loomp, waarin de auteurs o.a. schrijven:

²⁰ Zie De Roure D., Goble G. & Stevens, R. (2007). Designing the myExperiment Virtual Research Environment for the Social Sharing of Workflows. In: *e-Science 2007: Proceedings of the Third IEEE International Conference on e-Science and Grid Computing*, December 2007, Bangalore, India. Retrieved 10 September, 2009 from <http://eprints.ecs.soton.ac.uk/15838/>.

²¹ Zie: Integrated Content Environment (n.d.). Retrieved 10 September, 2009 from <http://ice.usg.edu.au/>

²² Zie: Sefton, P. & Downing, J. (2009). ICE-TheOREM: End to end semantically aware eResearch infrastructures for theses. Georgia Tech's Institutional Repository. Retrieved 10 September 2009 from <http://smartech.gatech.edu/bitstream/1853/28424/3/49-415-1-PB.pdf>

²³ Zie: eCrystals (n.d.). Retrieved 10 September 2009 from <http://ecrystals.chem.soton.ac.uk/>

²⁴ Zie: Liquid Publications (n.d.). Retrieved 10 September 2009 from <http://project.liquidpub.org/>

²⁵ Zie: Casati, F., Chiunchiglia, F. & Marchese, F. (2007). Liquid Publications: scientific Publications meet the Web. Changing the way scientific knowledge is produced, disseminated, evaluated and consumed. Retrieved 10 September 2009 from

<https://dev.liquidpub.org/svn/liquidpub/papers/deliverables/LiquidPub%20paper-latest.pdf>

"...Nowadays we can find interesting use cases with underlying business models in literature that are based on linked data contained in webpages. The BBC for example developed a system that utilizes automatic enrichment of content to increase the visiting time on their website. By providing links to information sources of their own website that are related to the currently shown website, a user does not need to search external sources for further information. *In general, it is more complex to identify the business value of publishing linked data on the Web, especially the advantage over search engines does not hold as an argument...*" [cursivering door M.F.]

Loomp wordt aangeprezen als "...a Web application for creating, managing, and publishing semantic data [...] that enables every Web user to produce semantically enriched content and to distribute it across various media easily..."

Naast DBpedia bevat het Web of Data op dit moment ook Geonames, US Census, EuroStat, MusicBrainz, BBC Programmes, Flickr, DBLP, PubMed, UniProt, FOAF, SIOC, OpenCyc, UMBEL, Virtual Observatories, en Yago.

Voorlopige conclusie: Linked Data is interessant maar nog lang geen main stream. Dit blijkt ook de tekst op de Linked Data [wiki](#) waar staat:

"...Continued growth of the Web of Data/Semantic Web is heavily dependent on four different aspects of the Linked Data ecosystem:

- the availability of vocabularies & ontologies,
- scalable and stable technologies,
- services which facilitate the use of linked data and also
- innovative application scenarios on top of the Semantic Web ..."

Naast het feit dat op de vier bovenstaande onderdelen nog flink wat werk verzet moet worden, is het grootste nadeel voor de gebruiker dat er op dit moment aparte browsers nodig zijn en dat de gebruiker de moeite moet nemen om iets te weten over het Semantic Web. Dat lijkt voorlopig nog een te hoge drempel.

DOAJ

KB heeft recent een contract gesloten om DOAJ in het E-Depot op te gaan nemen. Gebleken is dat een aantal nieuwe open acces tijdschriften experimenteren met data policies.

Andere ontwikkelingen

Sommige lokale vakgroepen en onderzoeksgroepen hebben eigen oplossingen voor compound documents gekozen. Zie bijv. <http://www.myexperiment.org/users/18> en <http://www.myexperiment.org/users/622> voor packages van Amsterdam resp. Twente.

In dit verband is het interessant te kijken naar wat iemand als Michael Nielsen heeft te zeggen:

"...What I will do instead is draw your attention to a striking difference between today's scientific publishing landscape, and the landscape of ten years ago. What's new today is the flourishing of an ecosystem of startups that are experimenting with new ways of communicating research, some radically different to conventional journals. Consider [Chemspider](#), the excellent online database of more than 20 million molecules, recently [acquired](#) by the Royal Society of Chemistry. Consider [Mendeley](#), a platform for managing, filtering and searching scientific papers, with [backing](#) from some of the people involved in Last.fm and Skype. Or consider startups like [SciVee \(YouTube for scientists\)](#), the [Public Library of Science](#), the [Journal of Visualized Experiments](#), vibrant community sites like [OpenWetWare](#) and the [Alzheimer Research Forum](#), and dozens more. And then there are companies like [Wordpress](#), [Friendfeed](#), and [Wikimedia](#), that weren't started with science in mind, but which are increasingly helping scientists communicate their research. This flourishing ecosystem is not too dissimilar from the sudden flourishing of online news services we saw over the period 2000 to 2005..."

[Mackenzie Smith](#) van MIT heeft ook een interessante visie in zijn artikel over Scientific Research communication.

Tenslotte is uit de veelheid van interessante ontwikkelingen nog te noemen het [Virtual Laboratory E-Science](#)

Bijlage 4 - Overzicht repositories

Instelling	IR platform	SURFshare afspraak versie
Universiteit Utrecht	Dspace	2.0
Fout! Verwijzingsbron niet gevonden. en ISIM	Dspace	2.0
Universiteit VU Amsterdam	Dspace	2.0
Universiteit Radboud Nijmegen	Dspace	2.0
Universiteit Amsterdam , AUP, KIT, Naturalis, HvA	Arno	3.0
Universiteit Maastricht	Arno	2.0
Universiteit Tilburg en NLDA	Arno voor artikelen N@tschool voor Lorenet	3.0 2.5
Universiteit Twente	Eprints	2.0
Universiteit Delft	Zelf ontwikkeld systeem obv Fedora	2.0
Unversiteit Wageningen	Zelf ontwikkeld systeem	2.0
Universiteit Erasmus Rotterdam	Zelf ontwikkeld systeem obv Dspace	3.0
Universiteit Groningen	Zelf ontwikkeld systeem	3.0
Hogeschool Saxion	Zelf ontwikkeld systeem	1.0
Hogeschool Fontys	Arno voor artikelen N@tschool voor Lorenet	3.0 2.5
Hogeschool Rotterdam	N@tschool voor Lorenet	2.5
KNAW	Eprints	2.0
CWI	Eprints (wordt Fedora)	2.0
HBO Kennisbank	SURFsharekit (=Fedora)	1.0
DANS	Zelf ontwikkeld systeem	N.v.t.
Inst. Social Studies	Fez	2.0
Nivel	?	?
Open Universiteit	Dspace	2.0
RIVM	Biomed Central (hosted)	N.v.t.
TU Eindhoven	Zelf ontwikkeld systeem	2.0

Bijlage 5 - De workshop verrijkte publicaties 25 november 2009

Als onderdeel van de bezinning op het thema Verrijkte Publicaties als onderdeel van het SURFshare Programma, organiseerde SURFfoundation een workshop.

1. Input voor de workshop

Voorafgaand aan de workshop ontvingen de deelnemers een concept versie van dit rapport. Hierin werd een schets gegeven van de gewenste situatie (doelstelling per eind 2011) en een concrete uitwerking om deze situatie te bereiken. Dit vormde de basis van een discussie aan de hand van een aantal stellingen over het doel en de weg waarlangs het doel bereikt zou kunnen worden.

Een passage uit het concept SURFshare Activiteitenplan 2010

"..De verbreding en verdiegingsstappen die in de volgende fase van SURFshare zullen worden genomen zijn:

- een kwantitatieve groei van verrijkte publicaties (streefgetal: 2000 stuks per eind 2011)
- vergroten van de bekendheid met verrijkte publicaties en de mogelijkheden die verrijkte publicaties voor onderzoekers bieden
- stimuleren van aanleveren van onderzoeksdata
- doorontwikkelen van infrastructuur die Verrijkte publicaties ondersteunt
- In kaart brengen van een workflow voor het maken, opslaan en verspreiden van verrijkte publicaties

Bestuderen van de mogelijkheden en ontwikkelen van mechanismen om de kwaliteit van verrijkte publicaties en alle componenten van een verrijkte publicatie te kunnen beoordelen.."

Voorstel voor concrete uitwerking (zoals ingebracht door Martin Feijen)

Als concrete doelstellingen zou gedacht kunnen worden aan een situatie per eind 2011 met de volgende concrete zaken:

Doel: alle HO instellingen die in SURFshare participeren hebben lokaal een platform (repository of vergelijkbaar platform) dat geschikt is voor de verwerking van eenvoudige VPs. Middel: RING activiteit van SURF.

Doel: Eerste helft van 2010 een heldere afspraak die het begrip VP voor een aantal disciplines (sociologie, economie, psychologie, rechten) helder uitwerkt en via richtlijnen voor opslag, metadatering en harvesting, definieert. Middel: WRM en WISH afspraken en richtlijnen plus support vanuit SURF (A-team)

Doel: Op basis van de VP afspraak worden in het beschikbare aanwezige arsenaal in Nederlandse archieven, onderzoeksgroepen, bibliotheken inventarisaties gemaakt van de basiscomponenten (datasets, artikelen, syntaxen, DAI's, PID's). Streefgetal is minimaal 20.000 VPs (voorzien van relaties tussen datasets, DAI's en PID's) per einde 2011.

Doel: Bouw van een zichtbare en aansprekende dienst die duidelijk de voordelen van VPs laat zien voor de verschillende stakeholders: onderzoekers, subsidiegevers, OZ managers. Middel: een bestaande (Narcis?) of nieuwe dienst die naar analogie van LOREnet de gebruiker in staat stelt VPs te lokaliseren, globaal in te zien en desgewenst te downloaden (of opvragen) voor hergebruik.

Doel: Transport van VPs vanuit lokale platforms naar de Dienst (showcase) en naar de archieven (KB, DANS, 3TU) is operationeel en geschikt voor productie. Middel: bestaande DIDL en ORE tools en afspraken verder operationaliseren op nationaal niveau. Afgeleid doel: efficiëntere harvesting foutcontrole voor de archieven en diensten.

Om de bovenstaande doelstellingen te bereiken is nodig:

- Focus: De activiteiten van HO instellingen en SURF in de komende twee jaar op het thema VPs worden met focus aangestuurd, uiteraard in samenhang met het thema Collaboratories. Dat betekent dat de twee werkpakketten voor deze twee thema's leidend zijn en dat de andere werkpakketten faciliterend, ondersteunend zijn.
- Projectplan: Het Australische ANDS projectplan zou prima als vertrekpunt kunnen dienen.
- Middelen: Financiering vanuit Prima of ander fonds.
- Commitment: De SURFshare community dient zich vooraf te committeren aan de doelstellingen.
- A team: Vanuit SURF zal support geboden moeten worden om het gat tussen voorhoede en anderen te dichten.
- SMART: Een zeer concrete en doelgerichte omschrijving van VPs, disciplines, doelen, aanpak, workflows, richtlijnen, taken, producten. Dit is onderdeel van het projectplan.
- Mindshift: In communicatie en aanpak de aandacht verleggen naar data als leidend thema. Publicaties alleen aandacht geven als onderdeel van VPs.
- Buzzwords: Inspiratie. Enthousiasme. Drive. Problem owner.
- Hergebruik: Niet alles zelf maken of bedenken maar zoveel mogelijk overnemen van elders wat bruikbaar is. Voorbeeld: de EsCiDoc toolset biedt waarschijnlijk erg veel nuttige tools.

2. Stellingen zoals gebruikt in de workshop

De hiervoor beschreven voorstellen voor doel en aanpak zijn gepresenteerd en besproken tijdens de workshop over Verrijkte Publicaties. Een groep van ongeveer dertig vertegenwoordigers van de SURFshare community (universiteitsbibliotheken, KNAW, KB, DANS, NBIC, SCP) heeft zich daarna gebogen over de volgende stellingen:

► Stelling 1

Het thema RM dient de komende twee jaar met energie en focus via een inspirerende showcase tot een concreet resultaat te gaan leiden; dit resultaat zal het SURFshare programma eind 2012 bij interne en externe stakeholders in positieve zin op de kaart zetten.

Reacties van de workshop deelnemers op stelling 1:

De term Research Maps wekt verwarring op. Het is beter vast te houden aan de term Verrijkte Publicaties.

Voordat we aan een showcase kunnen denken zijn er nog belangrijke infrastructurele zaken die voorrang moeten hebben.

- 1) Voorzieningen voor opslag van data, als onderdeel van de workflow van een onderzoeker, moeten met prioriteit worden onderzocht en, als dienst aan de onderzoeker worden aangeboden.
- 2) Het toevoegen van identifiers (PID, DAI) aan onderdelen van een VP is nog geen standaard procedure en moet daarom eerst als eerste activiteit aan de workflow worden toegevoegd na opslag van die onderdelen.
- 3) Het maken van landelijke afspraken of richtlijnen over opslag, verwerking, transport enz. van VPs op zodanige wijze dat dataproviders en serviceproviders samen tot werkende processen kunnen komen.

Deze infrastructurele basisvoorzieningen moeten als onderdeel van de lokale en van de landelijke infrastructuur worden ontwikkeld in onderlinge samenhang. Zonder deze voorzieningen is het ontwikkelen van diensten voor VPs niet goed mogelijk.

De term showcase wordt op verschillende manieren uitgelegd. Bedoeld is een operationele dienst, bijv. in de vorm van een portal vergelijkbaar met Narcis. Het is dus veel meer dan een demonstrator of een tijdelijk voorbeeld. Als we inderdaad kiezen voor deze aanpak dan zou het de

voorkeur hebben dat alle repositories daarin meedoen, onder lokale aansturing van de bibliotheek. Laat in de portal duidelijk de meerwaarde zien van VPs, bijv. door het zichtbaar maken van het nut van koppelingen (voorbeeld: een foto van een papyrus met de daarop gebaseerde artikelen) en door zichtbaar maken van mogelijkheden voor **hergebruik** van:

- een experiment
- onderdelen van (een) publicatie(s)
- data

Verschillende stakeholders moeten op een voor hen relevante manier de voordelen van VPs kunnen zien. Voor de onderzoeker die nu zelf een VP maakt in de vorm van een zelfgemaakte website moet uitgelegd worden wat de voordelen zijn van een VP op basis van een solide infrastructuur. De showcase portal moet geen eenmalig "speeltje" zijn maar een blijvende, structurele dienst zijn. Daarin moet het aspect duurzaamheid ook een rol spelen (link naar E-Depot?). Focus op vijf disciplines zou kunnen maar wellicht is focus op een aantal generieke onderzoeksmethodieken een goed alternatief.

► **Stelling 2**

De beschreven gewenste situatie is in lijn met de wensen van de SURFshare community.

Reacties van de workshop deelnemers op stelling 2:

De beschreven doelstelling en aanpak sluiten globaal aan bij de wensen maar met een aantal kanttekeningen (zie bij de afzonderlijke stellingen). Meer focus en inperking zijn zeker belangrijk. De beschreven aanpak sluit aan bij internationale ontwikkelingen. De vraag die resteert is: sluit het aan bij wat de onderzoekers willen?

Het is belangrijk om in zo vroeg mogelijk stadium in het onderzoeksproces te starten met data capture om later tot een VP te kunnen komen. We moeten aansluiten bij de methodieken en workflows van de onderzoeker en niets opleggen. Hier ligt een duidelijk raakvlak met het thema Collaboratories.

In de komende twee jaar moeten we niet alleen de alfa en gamma disciplines bedienen maar ook aandacht schenken aan de beta disciplines en hen actief betrekken bij het thema VPs.

► **Stelling 3**

De beschreven gewenste situatie, doelstellingen en resultaten is/zijn realistisch en haalbaar.

Reacties van de workshop deelnemers op stelling 3:

Kwaliteit is belangrijker dan kwantiteit. Een streefgetal is daarom geen doel op zich en moet weliswaar niet uit het oog verloren worden, maar moet niet leidend zijn in de aanpak en uitvoering. Als we bijv. in 2010 de nadruk leggen op infrastructuur en in 2011 op productie dan zou een getal van 5000 per einde 2011 een mooi streefgetal kunnen zijn.

Doel en middel lopen in de beschreven situatie en de weg daar naartoe door elkaar. Gaarne aanpassen.

Driver is wel degelijk relevant voor VPs. In Dnet zitten oplossingen voor compound docs en enhanced publications. Die zijn mogelijk bruikbaar en we moeten uitzoeken wanneer die Dnet versie beschikbaar gaat komen. Het lijkt er op dat ook in de komende jaren VPs in Dnet ondersteund zullen worden. Kortom: uitzoeken en contact houden met de DRIVER community.

► **Stelling 4**

De beschreven aanpak dient vervangen te worden door een andere, namelijk: De reden hiervoor is:

Reacties van de workshop deelnemers op stelling 4:

Het is belangrijk de resultaten van de tenders breed uit te dragen en te bekijken op welke wijze de resultaten opgeschaald kunnen worden. Wees daarbij helder over verwachtingen. Kijk naar de business case van VPs, beschrijf use cases, ontwikkel views voor de afzonderlijke stakeholders en laat daarmee de meerwaarde van VPs zien. Geef prioriteit aan het inventariseren en metadateren van data(collecties). SURF moet hierin het voortouw nemen en coördineren. Deze aanpak is beter dan het maken van een showcase portal met 20.000 VPs. Je zou wel kunnen denken aan een opzet

van een portal met 100 a 1000 hoogwaardige VPs die openbaar beschikbaar zijn. Maar daarnaast moet er via tenders doorgewerkt worden aan het maken van productierijpe tools voor het maken van VPs. Tevens moeten er afspraken komen over metadaten.
Denk ook aan subject portals en verrijkte content (bijv. weblectures).

► **Stelling 5**

Er zijn andere kansen die in het rapport niet aan de orde zijn gekomen, namelijk: Deze kansen moeten door Surf apart / als onderdeel van het thema Research Maps worden opgepakt.

Reacties van de workshop deelnemers op stelling 5:

Kijk niet alleen naar disciplines waar het relatief eenvoudig is om de meerwaarde van VPs aan te tonen (zoals archeologie, sociologie of bestuursrecht) maar kijk ook naar methodieken die door meerdere disciplines heen gebruikt worden (zoals het vergaren en verwerken van statistische data uit vragenlijsten). En kijk ook op dit vlak naar de relatie met collaboratories.

Bijlage 6 - Verslagen van de gesprekken met tenderprojectleiders

Verslag van gesprek Veteran Tapes

Gesproken met: Paula Witkamp en Maarten Hoogerwerf

Datum: 13 oktober 2009

Ervaringen vanuit project Veteran Tapes

Het Veteranen Instituut beheert ca. 1000 interviews (geluidsopnamen) met veteranen uit de periode WOII tot heden. Derig interviews worden nu in het project gebruikt met als doel het hergebruik van de bronnen te stimuleren. Onderzoekers uit zes verschillende disciplines werken met dit materiaal. Dit zal leiden tot zes verrijkte publicaties. Uitgave waarschijnlijk via AUP. Datasets worden bij DANS gedeponeerd. Er is een tool ontwikkeld (de zgn fragmentknipper) om audiofragmenten te koppelen aan de artikelen. Uitdagingen (problemen) zijn:

- Linken van fragment aan artikel; URN is niet voldoende; de tijdcodes van de fragmenten worden voor aanvullende identificatie gebruikt (ook in resolver ingebouwd);
- Metadata van artikel maar ook fragment, vooral technische metadata die nodig is voor gebruik van het audiofragment (file type, bitrate, lengte, etc.);
- Privacy issues; sommige uitspraken moeten om diverse redenen geanonimiseerd worden; dit levert een bewerkte dataset op die naast het origineel bestaat; welke versie deponeren?
- Iedere onderzoeker gebruikt voor zijn onderzoek eigen (audio)tools;
- Wie is eigenaar van de dataset? Vraag wordt relevant als onderzoeker migreert naar andere instelling;
- Het verrijken (via fragmentknipper) op zichzelf kost niet erg veel tijd; het anonimiseren kost wel veel tijd;
- Surfnet biedt nog geen streaming audio oplossing die kan werken met datasets in een archief of IR; daarom moet DANS een extern bedrijf inschakelen;
- ReM's worden met de hand gemaakt; voor 6 VP's geen probleem, maar bij grotere aantallen moet een geautomatiseerde oplossing beschikbaar komen.

Nota bene: onderzoeker krijgt geen credits voor deponeren noch voor verrijken. Kost alleen maar tijd. Animo is dus afwezig. NWO stelt wel eisen maar geen sancties. Universiteiten doen niets. VWS doet dat wel door 10% subsidie achter te houden tot na deponering. Archeologen moeten deponeren via wettelijke verplichting. Daardoor nu in DANS >7000 datasets (2000 groei per jaar), grotendeels archeologie.

Observaties over VP's in algemene zin

ORE is op zichzelf goed bruikbaar maar er is nog veel pionierswerk, finetuning, en pragmatisme nodig om tot bruikbare resultaten te komen die ook opgeschaald kunnen worden. Net als in de begintijd van OAI-PMH zijn nog veel aanvullende afspraken nodig en die kunnen alleen maar uit de praktische toepassing van ORE te voorschijn komen. Wat dat betreft is de ervaring binnen de SURF tenders nog mager: er zijn 7 tenderprojecten uitgevoerd per eind 2009. Ieder project ziet een deel van het totale probleemgebied en trekt daar zijn eigen, "lokale" conclusies uit.

Naast een behoefte aan meer praktijkervaring is het nodig om afspraken te maken tussen partijen. Uitgevers en / of bibliotheken (IR managers) moeten regels gaan opstellen voor aanlevering en structurering van VP's. Wederzijdse verantwoordelijkheden moeten vastgelegd worden tussen onderzoeker, UB, uitgever, archief, IR manager enz. Hoe minder de onderzoeker zelf hoeft te doen hoe beter. Daarom moet de workflow en werkverdeling beschreven worden.

Het gaat erom interesse te wekken voor VP's. Dat kan alleen door de ontwikkeling van diensten. Eigenlijk is een VP op zichzelf al een dienst! Onderzoekers willen best weten wat er met hun data

gebeurt – zijn er artikelen geschreven op basis van mijn data? Door wie, wanneer? Maar is mijn data dan correct gebruikt? Enz.

Infrastructuur kan grotendeels generiek gebruikt worden. Tools, data, metadata, workflow, diensten zijn meer discipline specifiek.

De manier waarop de relatie tussen artikel en dataset wordt gelegd heeft grote gevolgen. Hoe “dieper” de verrijkingen in het artikel gelinkt worden, hoe rijker de functionaliteit voor de gebruiker, maar hoe groter ook de complexiteit van de VP. JALC heeft hier al ervaring mee opgedaan. Het linken op metadata niveau tussen artikel en data is een stuk eenvoudiger.

In DARE lag de nadruk vooral op disseminatie aan de voorkant. Wat ontbreekt is een centraal loket aan de achterkant waar de onderzoeker heel laagdrempelig in een keer zijn data plus artikel kan aanleveren in ruil voor meerwaarde aan de voorkant. Of zou een CRIS zo’n loket kunnen zijn?

Aanvulling achteraf van Paula:

Privacy issues; sommige uitspraken moeten om diverse redenen geanonimiseerd worden; dit levert een bewerkte dataset op die naast het origineel bestaat; welke versie deponeren?

Wat Maarten zegt klopt, de gedeponeerde bestanden blijven altijd het eigendom van de deponerende in dit geval het Veteranen Instituut.

Je gaat dat denk ik in de toekomst toch meer krijgen dat er een mix ontstaat van eigen en andermans data. In dat geval levert dat een nieuwe dataset op.

In geval van Veteran tapes is het gebruik van ‘bewerkte dataset’ een beetje misleidend, het gaat om bewerkte fragmenten. Dit zijn geen datasets in de strikte zin van het woord maar delen van een dataset die al bij DANS gedeponerd is. Een bewerkte datasets, zou bijv. een geannoteerd interview kunnen zijn, dit is dan een nieuwe datasets die een relatie heeft (in de metadata vermeld) met een andere dataset. In het geval van een interview met piepjes, de referentie of de noot zal verwijzen naar het origineel (met PI en tijdcode), maar de verschijningsvorm/presentatievorm is die met piepjes. Dit is overigens ook nog een punt waar we het niet over gehad hebben, de status van de afgeleiden

Verslag van gesprek project Proefschriften Plus

Gesproken met: Martin Slabbertje

Datum: 7 oktober 2009

Samenvatting gesprek

Het CvB van de UU heeft recent als beleid bekend gemaakt dat medewerkers aangespoord moeten worden om hun publicaties (via Metis) te deponeren in het IR, waar die publicaties, indien toegestaan, publiek toegankelijk komen. Want: resultaten van met publiek geld gefinancierd onderzoek moeten publiek beschikbaar zijn. Het deponeren (en, desgewenst na een embargo-periode, openbaar maken) van proefschriften is verplicht gesteld. Het deponeren (en, desgewenst na een embargo-periode, openbaar maken) van scripties is een facultatieve dienst die de UB aanbiedt aan faculteiten. Er is nog geen beleid voor het deponeren van datasets.

Er loopt een apart project Podium olv Eric Sieverts waarmee de UU experimenteert met lokale opslag van datasets die niet in DANS of een ander data archief opgeslagen kunnen worden. Bijv. datasets van geowetenschappen.

Voor het projecten Proefschriften Plus gebruikt de UU Dspace versie 1.5.2, aangevuld met zelf gemaakt aanvullingen voor resource maps. Er zijn 500 promovendi uit alle sectoren benaderd maar slechts 20 van hen doen daadwerkelijk mee, vooral uit de Beta sector. Reden voor dit relatief geringe aantal is o.a. de open access eis (voor de sector rechten een echte showstopper), werklast en het feit dat het project eigenlijk net op het verkeerde moment (grote stress en drukte ivm promoveren) aanklopt.

Er wordt een Excel sheet gebruikt voor het aanleveren van metadata over de datasets door de promovendi. Daarna vindt een gesprek met hen plaats wat een visuele representatie van de VP

oplevert (samen met Igitur). Vervolgens voeren catalografen de metadata in Dspace in samen met een scan van de visuele representatie. Die leidt weer tot een resource map die straks gebruikt gaat worden in de nog te ontwikkelen showcase. De Vp's zullen niet geharvest worden, maar nu al is duidelijk dat er behoefte is aan aanvullende afspraken (richtlijnen?) over de toepassing van de ORE standaard.

Het hele proces van verwerven en invoeren van de vp's is erg tijdrovend: per proefschrift ongeveer 10 uur. Alles is nieuw, gebeurt met de hand, tools ontbreken nog, formulieren zijn er nog niet, de workflow is zeer gevarieerd omdat de datasets heel divers van aard zijn. Het lijkt erop dat er geen sprake zal kunnen zijn van een generieke aanpak. Eerder van een discipline specifieke aanpak.

Er is wel gekeken naar de ervaringen met het E-thesis project van Knowledge Exchange maar behalve het daar ontwikkelde semantisch model kan PP niet veel gebruiken. PP gaat veel verder en onderzoekt nieuw, onontgonnen terrein.

Hoe verder?

Op de vraag "hoe verder met VP's na 2009?" komt als antwoord dat een landelijke aanpak, gecoördineerd door SURF, betere perspectieven biedt dan opnieuw een aantal tenderprojecten. Daarbij zou o.a. aan de volgende zaken aandacht besteed moeten worden.

Organisatie

1. Welke datasets kunnen/moeten in welk data archief gedeponeerd worden en wat moet een universiteit (UB) lokaal doen?
2. In het verlengde van het vorige punt: welk archief is verantwoordelijk voor welke specifieke formats (datatype/applicatie) van datasets? Het is gevaarlijk de gebruiker hiermee op te zadelen.
3. Generieke aanpak versus discipline specifiek. Het lijkt er voorlopig op dat een generieke aanpak niet haalbaar is.
4. Voor proefschriften, maar waarschijnlijk ook onderzoek, is het zaak in zo vroeg mogelijk stadium de auteur rekening te laten houden met het feit dat de datasets gedeponeerd moeten worden en dat de auteur daartoe metadata moet aanleveren.
5. Issues rondom rechten op datasets (o.a. hergebruik) moeten opgelost worden.

Techniek

1. Zijn er systemen geschikt voor verwerking van datasets, zo ja welke, en wat dient er aanvullend nog een afspraken en tools bijgemaakt te worden en door wie?
2. Met name afspraken over het relateren van (delen) van datasets binnen een archief en tussen archieven onderling zeer relevant en nodig.
3. Er moeten tools, procedures, formulieren etc komen om het werk voor de auteur en de archiefbeheerder lichter te maken. Zonder die tools kosten VP's enorm veel tijd en dat is een groot risico voor de business case.
4. Hoe leg je de relatie tussen een VP en bijv. een VKC (virtueel kenniscentrum) zoals in UU?

Verslag van gesprek Dataplus

Gesproken met: Ellen Verbakel en Thomas Place

Datum: 15 oktober 2009

Ervaringen vanuit project Dataplus

De aanpak binnen Dataplus is voor een belangrijk deel gebaseerd op hergebruik van al beschikbaar materiaal: artikelen in repositories, datasets in archieven (DANS en GESIS), syntaxes op pc's van onderzoekers. Bestaande artikelen op EVS data zijn gedestilleerd uit de repository van de bibliotheek, maar daarnaast zijn er ook nog veel aanvullingen gedaan om te komen tot een complete EVS-repository en in het geval van NKO tot een complete NKO-repository. De ruwe data staat in het archief, de bewerkingen die op de data zijn uitgevoerd staan beschreven in de syntax en daaruit resulterende tabellen staan in het artikel. Dataplus VP's zijn DDI 3.0 compliant.

Demetaddata, het artikel (PDF) en DDI samen worden als DIDL package opgeslagen. Via een DIDL – ORE gateway is ook ORE uitvoer mogelijk. Maar ORE is voor het project op zich niet nodig. Het verrijgingsproces wordt ondersteund door een binnen het project ontwikkelde Enhanced Publication Editor. De onderzoeker gebruikt deze EPE voor het aanvullen van metadata met keywords, landen en jaartallen (keuzelijstjes), voor het toevoegen van de syntax en voor het kiezen van de juiste dataset uit en lijst van aanwezige datasets. Dit hele proces kost na enig oefenen niet meer dan 15 a 30 minuten werk. Als de syntax zoek is kan er niets toegevoegd worden.

Hoewel de syntax een relatie heeft met de gebruikte analytische / statische software is dat geen probleem. Een andere onderzoeker kan de syntax vrij eenvoudig hergebruiken ook al hanteert hij andere software. Onderzoekers zijn soms wel huiverig om de syntax op te sturen omdat zo iets voelt als "je bloot geven" en controle.

De Dataplus aanpak zou op zich kunnen werken in alle disciplines en situaties waar een onderzoeker werkt met experimentele data of survey data. Bijv: psychologen, sociologen, economen, politicologen. Datasets kunnen uit NL archieven komen maar ook uit internationale archieven. Waarschijnlijk is de internationale omvang wel groter, in ieder geval voor sociologie is dat het geval.

Observaties over VP's in algemene zin

Als het weinig extra moeite kost willen onderzoekers wel meewerken aan verrijken. Maar er is (veel) meer nodig. Persoonlijke motivatie: het vergroten van de zichtbaarheid van je werk. Een referentie naar je artikel levert credits op. Zo zou het ook moeten zijn met een referentie naar de dataset of syntax. Intrinsieke motivatie: het dienen van het belang van de wetenschap in algemene zin.

Wat ook (of in eerste instantie wel het meest) zou helpen is uitgevers die de eis stellen dat de dataset en syntax meegeleverd moeten worden. NWO en andere subsidiegevers zouden dat ook kunnen eisen. Universiteiten zullen zo'n eis niet snel aan hun onderzoekers opleggen, want dat zou ten koste gaan van hun onderzoek- en onderwijstaken. CvB en universitaire management zou wel kunnen uitstralen en uitdragen dat het normaal is en daarom van een onderzoeker verwacht wordt dat hij / zij de dataset en syntax deponereert en daarmee hergebruik via VP's bevordert. Harde dwang werkt niet, een voorbeeld stellen wel.

2010 en verder?

SURF moet absoluut wel doorgaan met het thema VP's. Het zal een kwestie van lange adem zijn maar het is zeer zeker de moeite waard en het zal ook wel gaan lukken. Er wordt nu toch al redelijk vaak door onderzoekers over VP's gesproken. Het begint langzaam wat bekender te worden. Interessant zou zijn om in dat kader te kijken naar het aantal journals met een data policy. Waarschijnlijk zit daar groei in maar hoe veel?

De Dataplus aanpak zou opgeschaald kunnen worden. Maar belangrijker is dat SURF initiatieven ontwikkeld om de voorbeeldfunctie en intrinsieke motivatie bij onderzoekers onder de aandacht te brengen.

Tools als de EPE van Dataplus en de DVN tool (open source) zouden beschikbaar gesteld moeten worden aan alle onderzoekers, en doorontwikkeld moeten worden zodat ook andere disciplines er gebruik van kunnen maken, maar dan specifiek voor hun situatie.

De grootste voorkeur gaat in Tilburg uit naar een initiatief vanuit SURF om een landelijk project uit te voeren met de gehele SURFshare community gericht op quick wins op het thema VP's. Concreet: begin met de beschikbare datasets in archieven (inter)nationaal en kijk ook naar datasets op het web, oa. Via journals met een data policy. Zoek daar de artikelen bij, leg de relatie en maak er DDI compliant DIDL en ORE packages van. Gebruik de Dataplus en DVN tools via lokale of hosted installatie. Realiseer op die manier concrete, zichtbare meerwaarde voor de onderzoeker, bijv. in de vorm van publicatielijsten waarin ook de datasets vermeld zijn.

Een andere optie is om wederom een ronde tenderprojecten op te starten waarin de nadruk nu moet liggen op demo projecten die aantonen wat de meerwaarde is van ORE en VP's voor specifieke disciplines, bijv. economie, medicijnen.

Een combinatie van een breed landelijke (quick wins) project en een diepte demo project (bijv. economie) is natuurlijk ook mogelijk.

Let ook op de relatie met onderwijs. VP's en datasets hebben ook nut voor het onderwijs. Dit punt zou uitgebuit kunnen worden in het kader van het motiveren van onderzoekers om publicaties te verkrijgen.

JALC

In het kader van dit Rapport over VPs heeft Martin Feijen in oktober 2009 een projectvergadering van JALC bijgewoond. Van deze vergadering is voor dit rapport geen afzonderlijk verslag gemaakt. De observaties zijn uiteraard wel verwerkt in het rapport.

Verslag van gesprek Escape

Gesproken met: Esther Hoorn en Henk Ellermann

Datum: 27 oktober 2009

Ervaringen vanuit project Escape

Succesfactoren zijn:

- gebruik van open source software (zoals de Escape tool);
- maak het discipline specifiek; vooral alfa en gamma zijn kansrijk (Rechten, Archeologie, Sociologie, Economie, Erfgoed);
- Geef uitleg over auteursrechtelijke zaken;
- Focus op die onderzoeksgebieden waar de opdrachtgever eist dat datasets bewaard en bruikbaar blijven. Voorbeeld: Bestuursrecht;
- Neem een voorbeeld aan Wikipedia: mensen willen wel een bijdrage leveren als die maar niet te groot is.

Observaties over VP's in algemene zin

De term Verrijkte Publicaties dient vervangen te worden door iets beters. Het is nog te veel gelinkt aan het oude denken vanuit de publicatie.

De techniek is er wel, de organisatie nog niet. Er moet meer aandacht zijn voor de regulering, eigenaarschap, afgeleide rechten.

Er lijkt (ook binnen SURF) sprake te zijn van een merkwaardig spanningsveld tussen enerzijds de wetenschappelijke valorisatie (artikelen) en anderzijds de maatschappelijke valorisatie (datasets). Onderzoekers hebben er belang bij dat hun data zichtbaar en herbruikbaar zijn i.v.m. het binnen halen van 3^e geldstroom financiering. Dit raakt de business case van VP's. In het verlengde hiervan loopt de discussie over kwaliteit en impact. Maatschappelijke valorisatie is overigens voor het HBO erg belangrijk.

RDF is in opkomst en wordt steeds belangrijker. Voordelen beginnen nu ook zichtbaar te worden, zoals bijv. blijkt uit een experiment met RDF metadata van OCLC waarmee de gebruiker relaties kan zien tussen (samenwerkende) auteurs of relaties tussen onderwerpsgebieden (al dan niet grafisch weergegeven).

Linked data is een tool die open access bevordert. Daarmee wordt voorkomen dat datasets achter het hekje van commerciële uitgeverij verdwijnen (bijv. Cell).

Onderzoekers kijken nu naar commerciële uitgevers maar niet naar de UB. Het is hard nodig dat de UB zich zichtbaar en herkenbaar gaat opstellen als dienstverlener voor de onderzoeker.

Praktische problemen:

- Identifiers. DAI is er weliswaar maar de echte implementatie vergt nog heel veel werk. Hetzelfde geldt voor identificatie van documenten. De door SURF gemaakte standaardlijst voor documenttypen is erg waardevol.
- De werelden van repositories en Metis zijn nog te veel gescheiden.
- Binnen de RUG en wellicht ook elders bestaat een cultuur van : elk probleem zijn eigen systeem. Er is te weinig oog voor en draagvlak voor generieke infrastructuur. Daardoor kan de UB ook moeilijk als dienstverlener bij de onderzoeker succesvol zijn.

2010 en verder?

Allerbelangrijkste thema: infrastructuur bouwen.

1. identificatie goed regelen en via voorbeelden de voordelen laten zien;
2. RDF als standaard voor datamodel gaan gebruiken en met simpele vocabulaires en opwerken van metadata meerwaarde gaan creëren, bijv. door samenhang tussen data-entiteiten (dataset, artikel, auteur etc) onderling zichtbaar te maken;
3. verbijzonder naar discipline.

Hoe? Via een sterke voorhoede (en dan ontkom je helaas bijna niet aan tenders). Narcis als voorbeeld en drijvende kracht verder uitbouwen. SURF kan bibliotheken die niet zo snel mee kunnen in het tempo van de trekkers helpen.

Doel? Eind 2011 "semantisch" browsen in een rijke Narcis dienst (via standaard browsers).