

Een onderzoek naar de Open Data van Nederlandse (G)LAM instellingen



OPEN
GLAM

Eindverslag (revisie)

Datum: 07-03-2016

Vak: Integratieproject 2016

Aantal woorden: 16.876

Docenten: J. Kamps en H. Voorbij

Studenten: Annejet Heijboer, Tim van Laar, Sabine de Lange, Tineke van der Meer, Jan Willem Nieuwold, Tara Rijkers, Luciën Tull, Marco Roling, Arthur Vadimovich Teutler en Esmee van de Zand.

Contact: tineke.vandermeer@student.uva.nl

Voorwoord

Voor u ligt een onderzoek naar de open data van Nederlandse erfgoedinstellingen. Dit onderzoek is geschreven in het kader van het integratieproject binnen de master Culturele Informatiewetenschap aan de Universiteit van Amsterdam. Tien deelnemende studenten hebben dit onderzoek uitgevoerd in de maand januari 2016. Dit onderzoek is begeleid door de docenten Henk Voorbij en Jaap Kamps.

Gezamenlijk is een onderzoeksvraag opgesteld met daaronder drie subvragen, en vervolgens zijn de studenten opgedeeld in kleinere groepen. Drie groepen hebben elk een subvraag op zich genomen en één groep heeft zich gericht op het literatuuronderzoek, als volgt:

- Het uitvoeren van onderzoek en het schrijven van het verslag over Instelling en Aanbieder is gedaan door Marco, Tineke en Arthur;
- Het uitvoeren van onderzoek en het schrijven het verslag over Gebruik: GLAMetrics is gedaan door Jan Willem en Lucien;
- Het uitvoeren van onderzoek en het schrijven van het verslag over Gebruik en Gebruikers is gedaan door Tim, Sabine en Esmee;
- Het verzamelen van achtergrond literatuur, het schrijven van de inleiding + het theoretisch kader en het redigeren van het eindverslag is gedaan door Annejet en Tara.

Wij wensen u veel leesplezier.

Inhoudsopgave

Voorwoord	2
1. Inleiding	4
1.1 Aanleiding: open cultuurdata	4
1.2 Doel onderzoek	5
1.3 Probleemstelling	5
1.4 Leeswijzer	6
2. Theoretisch kader	7
3. Instelling & Aanbieder	11
3.1 Inleiding	11
3.2 Analyse en verrijking van de onderzoeksdata	12
3.3 Analyse van adoptie van open data	13
3.4 Hypothesen, statistische analyse en bevindingen	16
3.5 Korte samenvatting en conclusies	20
4. Gebruik: GLAMetrics	22
4.1. Inleiding	22
4.2 Werkwijze en onderzoeksvraag	22
4.3 Meetinstrumenten voor webstatistieken van Wikipedia	23
4.3.1 <i>BaGLAMa 2</i>	23
4.3.2. <i>GLAMorous</i>	24
4.4 Onderzoeksmethode	25
4.5 Meetresultaten voor de Nederlandse aanbieders op Wikimedia	26
4.5.1. <i>Bevindingen op basis van het GLAMorous-onderzoek</i>	27
4.5.2. <i>Bevindingen op basis van het BaGLAMa 2 onderzoek</i>	28
4.6. Conclusie	30
5. Gebruik & Gebruikers	32
5.1 Inleiding	32
5.2 Bevindingen	33
5.2 Conclusie	38
6. Conclusie	40
Literatuurlijst	43
Bijlagen	44

1. Inleiding

Monumenten, vondsten, opgravingen, nationaal beschermde voorwerpen en verzamelingen, archieven, boeken, brieven, foto's, film en andersoortig cultureel erfgoed, dikwijls met een grote cultuurhistorische en wetenschappelijke betekenis. Het is de representatie van onze cultuur en maakt ons bewust van onze geschiedenis. Het is daarom van belang dat dit zorgvuldig bewaard wordt in bijvoorbeeld musea, wetenschappelijke instellingen, archieven en bibliotheken. Het cultureel erfgoed wordt opgeslagen en beheerd in deze instellingen waar het tevens toegankelijk wordt gemaakt voor onderzoekers en het grote publiek. Deze culturele erfgoedinstellingen zijn de bewaarders en bewakers van onze gemeenschappelijke cultuur en kennis en dienen daarmee een publiek belang.

Om de zichtbaarheid van het erfgoed te verhogen en ook het origineel beter te kunnen behouden en beheren, zijn grote delen van archieven en collecties inmiddels gedigitaliseerd en voor een steeds groter publiek online toegankelijk gemaakt. Fragiele stukken die fysiek eerder niet toegankelijk waren vanwege hun conditie kunnen nu digitaal worden ingezien. Erfgoedobjecten die in depot zijn opgeslagen vanwege gebrek aan expositiemogelijkheden kunnen toch digitaal getoond worden. Ook wordt meer publiek bereikt omdat men niet meer fysiek naar de bewaringsplaats hoeft toe te reizen om de informatie en afbeeldingen van het erfgoedobject te kunnen bekijken. Door erfgoedobjecten te digitaliseren en de afbeeldingen en beschrijvingen online te publiceren en te delen wordt meerwaarde gecreëerd voor archieven en collecties en daarmee het publieke belang gediend.

1.1 Aanleiding: open cultuurdata

Met het digitaliseren van het cultureel erfgoed heeft er een verschuiving plaatsgevonden naar open data cultuur: het digitaal beschikbaar maken van representaties van collectiestukken en of kennis en informatie van culturele instellingen en initiatieven over hun collecties, activiteiten en organisatie. De zogeheten open data is dus data die door een ieder vrij gebruikt, hergebruikt en verspreid mag worden¹. De G8 stelde in 2013 het International Open Data Charter op waarin overheden zich verplichten hun data open te stellen². Dit werd vrijwel door de hele wereld onderschreven. Jaarlijks wordt door de Open Knowledge Foundation een ranglijst gepubliceerd van landen en hun status met betrekking tot open data. In 2015 stond Nederland op deze lijst op nummer 16, Groot-Brittannië op nummer 1. De plaats op de ranglijst wordt bepaald aan de hand van het aantal beschikbare datasets over een vast aantal onderwerpen: cultureel, wetenschap, financieel, statistiek, weer, milieu en transport. Volgens OpenDataNederland zijn er momenteel 617 open datasets binnen Nederland, deels met een erfgoedkarakter.

Op initiatief van de Open Knowledge Foundation is OpenGLAM (Galleries, Libraries, Archives and Museums) geïntroduceerd: een initiatief dat culturele instellingen ondersteunt in het openstellen van hun data door middel van workshops, documentatie, begeleiding en netwerken. Daarnaast bestaat het Nederlandse initiatief Open Cultuur Data (OCD), die tracht de Nederlandse culturele sector sterk te verankeren in de internationale open databeweging. Momenteel heeft OCD ruim 60 datasets van 37 instellingen geïnventariseerd³.

Door het beschikbaar stellen van digitale erfgoedobjecten en hun beschrijvingen ontstaan nieuwe mogelijkheden om deze data her te gebruiken: hieruit kunnen nieuwe toepassingen ontstaan, websites worden gemaakt en nieuwe manieren worden gecreëerd om te participeren in kunst en cultuur. Een goed voorbeeld hiervan is het besluit van het Rijksmuseum om de digitale afbeeldingen van het grootste deel van hun collectie in een hoge resolutie aan iedereen beschikbaar te stellen voor vrij gebruik in commerciële

¹ Zie: <http://www.opencultuurdata.nl/over/>

² Zie: <http://www.frankwatching.com/archive/2015/11/24/open-data-waarom-we-het-belgische-voorbeeld-moeten-gaan-volgen/#more-363770>

³ Zie: <https://opendatanederland.org>

toepassingen en producten. De zichtbaarheid van deze erfgoedobjecten en van het Rijksmuseum wordt hierdoor enorm vergroot. In de praktijk blijken gedigitaliseerde erfgoedobjecten en beschrijvingen vaak moeilijk vindbaar, matig toegankelijk en vanwege rechten beperkt bruikbaar. Dat terwijl het gedigitaliseerd cultureel erfgoed in de huidige informatiemaatschappij juist toegankelijker zou moeten zijn dan ooit. Maar is dat ook zo? En welke aspecten spelen dan een rol in het aanbieden van open data en het gebruik ervan?

1.2 Doel onderzoek

Tussen 2014 en 2016 heeft OpenGLAM een onderzoek gedaan onder erfgoedinstellingen wereldwijd. Het doel van het onderzoek was het meten van de staat van voortgang van OpenGLAM in deelnemende landen, het informeren van de GLAM community, het identificeren van potentiële partners, het promoten van OpenGLAM en het onderzoek als lobby instrument in te zetten voor activiteiten in het voordeel van OpenGLAM en het openstellen van kennis. Onder leiding van Beat Estermann is uitgebreid onderzoek gedaan onder erfgoedinstellingen en is een paper geschreven waarin een vergelijking is gemaakt tussen de vier deelnemende landen, waaronder Nederland. Van een specifiek Nederlandse verslag op basis van deze onderzoeksgegevens is echter nog geen sprake geweest tot nu toe, omdat men niet over voldoende middelen beschikt.

In het huidige onderzoek uitgevoerd door masterstudenten aan de Universiteit van Amsterdam is getracht een bijdrage te leveren aan het in kaart brengen van de stand van zaken ten aanzien van open data onder Nederlandse cultureel erfgoedinstellingen. Dit is gedaan door verschillende facetten van de open data cultuur nader te onderzoeken en te belichten. Er is gebruik gemaakt van de beschikbare onderzoeksdata voor Nederland uit het OpenGLAM onderzoek van Estermann en de verkenning van open data via het project OpenCultuurData. De usability van GLAMetrics tools is onderzocht om gebruik van open data te meten. Tevens is er een vragenlijst uitgezet onder erfgoedinstellingen met vragen over gebruikers en gebruik van open data. Om het onderzoek af te bakenen is gekozen voor de erfgoedsectoren archieven en musea, omdat er is geconstateerd dat in de gebruikte bronnen een gebrek aan data van en over bibliotheken aanwezig is.

1.3 Probleemstelling

Tijdens dit onderzoek is de volgende onderzoeksvraag gehanteerd: ‘Welke aspecten spelen een rol bij het aanbieden van open data van Nederlandse erfgoedinstellingen op dit moment en in de nabije toekomst en wat is de impact van het gebruik van open data?’. Om een antwoord op deze vraag te kunnen geven, zijn de volgende deelvragen geformuleerd:

1. Wat zijn de karakteristieken van erfgoedinstellingen die voorlopen met het beschikbaar stellen van open data?
2. Welke erfgoedinstellingen stellen op Wikimedia Commons open content en data beschikbaar voor (her)gebruik en hoe onderscheiden de voorlopers op dit terrein zich van de rest?
3. Wie zijn de gebruikers van open data, wat zijn de gebruikstoepassingen en hoe wordt gebruik van open data gemeten?

Om antwoord te kunnen geven op de eerste deelvraag zijn erfgoedinstellingen in Nederland het gebied van open data nader onderzocht. Hiervoor zijn beschikbare gegevens uit het OpenGLAM onderzoek van Estermann over open data in Europa als uitgangspunt genomen. Naast de bepaling van de grootte van de instelling en de adoptie van open data is in de dataset gekeken welke variabelen bij adoptie van open data een rol spelen.

Voor de beantwoording van de tweede deelvraag is het gebruik en hergebruik van open data van erfgoedinstellingen op Wikimedia gemeten. Uitgangspunt was de groep erfgoedinstellingen op het

OpenCultuurData platform waarvan bekend is dat deze open data beschikbaar hebben gesteld, hoewel niet allen aan Wikimedia. Door alleen Wikimedia te bestuderen is slechts een deel van het open data aanbod afgedekt, maar dit is toch gedaan als case study om te zien of impact van open data meetbaar is en wat daarmee kan worden onderzocht. Er is gekeken hoe open data van erfgoedinstellingen op Wikipedia pagina's terechtkomen en of beschikbare GLAMetrics tools bruikbaar zijn voor het meten van open data en het eventueel bepalen van de mate van adoptie van open data door erfgoedinstellingen.

Voor de beantwoording van deelvraag drie is een vragenlijst opgesteld en per e-mail aangeboden aan de groep erfgoedinstellingen op de OCD website, waarvan bekend is dat deze open data beschikbaar hebben gesteld. Dit om meer inzicht te krijgen in de gebruikers van open data, het gebruik en de gebruikstoepassingen. De respons is in de verslaglegging verbonden aan de oorspronkelijke deelvraag, welke als leidraad gold tijdens de uitvoering van het onderzoek.

1.4 Leeswijzer

Het verslag is als volgt opgebouwd: het volgende hoofdstuk is het theoretisch kader waarin de belangrijkste aspecten, definities en conceptuele modellen rondom open data in kaart zijn gebracht. In hoofdstuk 3, 4 en 5 is het onderzoek per deelvraag inzichtelijk gemaakt en zijn de belangrijkste conclusies uiteengezet. Hoofdstuk 6 bestaat uit de eindconclusie waarin antwoord op de hoofdvraag is gegeven en de belangrijkste aanbevelingen zijn opgenomen. Het onderzoek sluit af met een literatuurlijst en de bijlagen.

2. Theoretisch kader

Het theoretisch kader van dit onderzoek vindt zijn basis grotendeels in eerder genoemde onderzoek van Beat Estermann (2014) naar culturele erfgoedinstellingen uit Finland, Polen, Zwitserland en Nederland. In zijn onderzoek “Diffusion of Open Data and Crowdsourcing among Heritage Institutions” wijst Estermann op een gebrek aan onderzoek naar de trends op het gebied van adoptie van open data in Europa (2). Want hoewel open data momenteel wereldwijd op de agenda staat, is er vandaag de dag nog groot gebrek aan onderzoek binnen Europa.

Dit onderzoek tracht een bijdrage te leveren aan het theoretisch debat over open data door de huidige situatie van open data op het gebied van Nederlandse culturele erfgoedinstellingen in kaart te brengen. Het eerder verrichte onderzoek van Estermann geldt hierin als basis. Daarnaast is ook enige aandacht besteed aan meer algemene literatuur over open data en de adoptie van open data.

Culturele erfgoedinstellingen

Dit onderzoek volgt Estermann in de definitie van erfgoedinstellingen, welke de definitie van het ENUMERATE project volgt: “the cultural heritage domain consists of “memory institutions”, such as museums, libraries, archives and records offices, audio-visual and film archives, organizations with curatorial care for monuments, sites and the historic environment, as well as hybrid types of organizations” (Estermann 2).

Open data

De open data beweging bestaat al zo'n 50 jaar en vindt zijn oorsprong binnen het academisch circuit. Een wereldwijde doorbraak werd bereikt toen de Verenigde Staten en Groot-Britannië het Open Overheid Data beleid aannamen om transparantie, participatie en samenwerking onder politici, autoriteiten en burgers te stimuleren. Volgens de Sunlight Foundation wordt data als ‘open’ gezien als deze kosteloos kan worden hergebruikt, aangepast en verspreid door een ieder die dat wil: “In order to facilitate re-use, the data need to be made available in a machine readable format, i.e. as structured data. Typically, open data or content that is subject to copyright protection is made available under a free copyright license, which allows users to freely modify and to re-distribute a work” (Estermann 3). Dit onderzoek volgt de eigenschappen van open data die door de Sunlight Foundation zijn vastgesteld en verstaat onder ‘open data’ zowel de metadata als open content van gedigitaliseerd erfgoed. In het Open Data Handbook worden de eigenschappen waarover open data beschikt verder toegelicht:

- Beschikbaarheid en Toegankelijkheid: De data moet in zijn geheel beschikbaar zijn, en voor niet meer dan een redelijke productieprijs, bij voorkeur door middel van downloaden via het internet. De data moet ook beschikbaar zijn in een handige en modificeerbare vorm.
- Hergebruik en verspreiding: de data moet aangeboden worden onder voorwaarden die hergebruik en verspreiding toestaan, daarbij ook het samenvoegen met andere datasets inbegrepen.
- Universele deelname: iedereen moet kunnen gebruiken, hergebruiken en verspreiden - er moet geen discriminatie bestaan tegen velden van een bepaalde onderneming, of tegen personen of groepen.

In het beschikbaar en toegankelijk maken van de data is interoperabiliteit van groot belang om verschillende datasets te kunnen combineren. Wanneer het gaat om het bouwen van grote, complexe systemen waarin verschillende datasets gecombineerd kunnen worden, is interoperabiliteit zelfs vereist om openheid te kunnen realiseren.

Adoptie van open data

Het aantal culturele erfgoedinstellingen dat open data voor hergebruik aanbiedt groeit hard. In de praktijk is het helaas zo dat veel instellingen enorm worstelen met de (meer)waarde van hun aangeboden data. In de meeste gevallen is dit te herleiden tot de cultuur van de instelling. De cultuur is bepalend voor de mate waarin een culturele erfgoedinstelling bereid is om open data te omarmen (adoptie). Bovendien is er vanwege de organisatiecultuur totaal geen of heel weinig zicht op de eindverantwoordelijke voor de open data initiatieven.

Estermann deed in 2012 al onderzoek naar culturele erfgoedinstellingen in Zwitserland, om op deze wijze een instrument te ontwikkelen die de adoptiegraad van open data beleid en crowdsourcing toepassingen kon meten. Hiermee legde Estermann daarnaast ook de voornaamste uitdagingen en drijvende krachten bloot in de verspreiding van open data. In “Diffusion of Open Data and Crowdsourcing among Heritage Institutions” voorspelt Estermann dat er de komende vijf jaren sprongen binnen de huidige digitaliseringsactiviteiten zichtbaar zullen worden, waar nieuwe mogelijkheden uit kunnen voortvloeien (1). Volgens Estermann speelt de grootte van een instelling een positieve rol in de adoptie van open data. De mate van vrijwilligers zijn hierin van negatieve invloed: “Institutions whose workforce counts only volunteers are less likely to adopt open data than other institutions” (11). De organisatiecultuur (publiek, privaat nonprofit of commercieel) speelt volgens Estermann geen rol in de adoptie van open data, met uitzondering van instellingen met een gemixte samenstelling, welke negatief van invloed is. Daarnaast speelt de betrokkenheid van de instelling bij Wikipedia een positieve rol. Estermann stelt daarover echter dat “The cause-effect-relationship is unclear: for example, are they involved in Wikipedia because it’s part of their social media strategy? Or are they embracing social media, because they are involved in Wikipedia?” (11).

Een andere belangrijk aspect in het (her)gebruik van open data is het auteursrecht, wat duidelijk in de randvoorwaarden verwerkt dient te zijn. Estermann stelt hierover dat het auteursrecht geen limiterende factor is in het openstellen van de data: “Regarding the copyright status of heritage material, it appears that for all the object types, more than half of the objects are either in the public domain or the copyright is controlled by the institution. Thus, it appears that across for the average institution, copyright is not the limiting factor so far when it comes to opening up content” (18). Estermann meent dat het auteursrecht meer een beperking zal worden in het vrijgeven van open content.

Het positief aspect van licentievrije open data is dat organisaties en burgers met behulp van de culturele data meer data kunnen verzamelen om hiermee nieuwe technologische toepassingen te realiseren. Er zijn meer mogelijkheden met open data doordat er steeds nieuwe datasets worden ontsloten op verschillende platformen. Een van deze platformen, die in dit onderzoek daarom wordt geanalyseerd, is Wikimedia Commons: een meertalige website die als repository dient voor beeld-, video- en geluidsbestanden. Deze website is een geschikte locatie voor erfgoedinstellingen om hun open content en open data mondiaal beschikbaar te stellen voor het vrije (her)gebruik. Estermann stelt dat instellingen wiens personeel betrokken is bij Wikimedia Commons 35% meer kans hebben om open content strategie te adopteren (11).

Belemmeringen binnen open data van erfgoedinstellingen

In “Opening Access to Collections: the Making and Using of Open Digitised Cultural Content” onderzoekt Melissa Terras de relatie tussen de Open Access beweging en de huidige debatten over auteursrecht en beschikbaarheid van cultureel erfgoed. Terras stelt dat er nog een lange weg te gaan is voor het een algemene norm is dat erfgoedinstellingen hun collecties open stellen. Een grote barrière wordt volgens Terras gevonden in het idee van wat het voor een instelling betekent om ‘open’ te zijn: “Barriers remain, mostly in the conception of what it means for the institution - in both a financial and reputational sense - to behave in an open manner” (23). Terras wijst hierin op de senior individueel die niet perse deel uitmaakt van de digitale

cultuur maar wel de financiële keuzes maakt. Zij moeten volgens Terras meer kennis maken met de OpenGLAM beweging.

Daarnaast wijst Terras op zogenaamde 'Orphan Works' binnen de GLAM sector, welke in Groot-Brittannië bijvoorbeeld al 50 procent van de archiefcollecties beslaan: "creative works or performances that are subject to copyright - like a diary, photograph, film or piece of music - for which one or more of the right holders is either unknown or cannot be found" (24). Door ingewikkelde wet- en regelgeving ontstaat hier een probleem door het 'opsluiten' van 'Orphan' content: "with little or no prospect of these items ever making a meaningful contribution to a knowledge economy without potentially complex and costly 'due diligence' processes" (24).

Tegengeluid

Een kritisch tegengeluid in de bewering dat open data 'beschikbaar' en 'toegankelijk' is voor een ieder, is afkomstig van Michael Gurstein. In "Open Data: Empowering the Empowered or Effective Data Use for Everyone?" stelt Gurstein dat open data niet *iedereen* volmacht geeft. Sterker nog, het impliceert volgens Gurstein juist het tegenovergestelde: "open data" empowers those with access to the basic infrastructure and the background knowledge and skills to make use of the data for specific ends" (1). Gurstein wijst op het feit dat open data met name de rijkere mens ten goede komt: "These above mentioned resources are more likely to be found among those who already overall have access to and the resources for making effective use of digitally available information. This would then suggest that a primary impact of "open data" may be to further empower and enrich the already empowered and the well provided for. On the other hand, those most in need of the benefits of such new developments may find themselves out of luck" (1).

Volgens Gurstein zouden er een aantal overwegingen in de processen en elementen van open data moeten worden opgenomen om de effectiviteit van het gebruik te verzekeren. Hieronder vallen volgens Gurstein kosten en beschikbaarheid van internettoegang, de taal van de data, technische en professionele vereisten voor het interpreteren en gebruiken van data en de beschikbaarheid van training in datagebruik en -visualisatie. Gurstein introduceert een model om effectief gebruik van open data mogelijk te maken en meent dat van elk van de zeven componenten in ieder geval gedeeltelijk aanwezig moeten zijn om geen gebruikers buiten te sluiten: "Some component of each needs to be in place or large numbers of those who might otherwise make use of open data to improve their lives and particularly the poor and marginalized will be excluded from making "effective use" of open data" (3). Het model bestaat uit de volgende componenten:

Michael Gurstein - A model for effective data use (2011)
<p>1. Internet — having an available telecommunications/Internet access service infrastructure sufficient to support making the data available to all users. Issues here would include:</p> <ul style="list-style-type: none"> a. the affordability of Internet access — a major issue for many particularly in the developing world. b. the availability of sufficient bandwidth for the range of uses to which the data might be effectively put, for example, whether the data access has been designed on the basis that for example, broadband is necessary for the use of the data being made available. c. the accessibility of the network, for example, where access to the network or to connectivity is restricted for political or other reasons. d. physical accessibility/usability of access sites as for example for the physically disabled.
<p>2. Computers and software — having access to machines/computers/software to access and process the available data and machines that are sufficiently powerful to do various analyses; having sufficient time on the equipment to do the analyses (many people need to share computers); knowledge of how to operate the equipment sufficient to access and analyze the data and so on. Does the use of the data require more powerful (and expensive) computers or software than might be generally available, for example?</p>
<p>3. Computer/software skills — having sufficient knowledge/skill to use the software required for the analyses/making the mashups/doing crosstabs, etc. Techies know how to do visualization, university trained persons and professionals know how to use the analytical software but ordinary community people might not know how to do either and getting that expertise/support might be either difficult or expensive or both.</p>
<p>4. Content and formatting — having the data available in a format (language, coding for display, appropriate geo-coding) to allow for effective use at a variety of levels of linguistic and computer literacy. What are the language, computer literacy, data analytic literacy levels that are required for an effective use of the "open data"? Does the use of the data presume that it is being used by a professional and are there means through which those professionals might be available to those who can't afford expensive fees?</p>
<p>5. Interpretation/Sense making — sufficient knowledge and skill to see what data uses make sense (and which don't) and to add local value (interpretation and sense making); being able to identify the worthwhile information and to figure out how to put the data into the right format or context so that what might otherwise be numbers on a page becomes something that can change people's lives.</p>
<p>6. Advocacy — having supportive individual or community resources sufficient for translating data into activities for local benefit. Availability of skills and local resources, community infrastructures, training, the means for advocacy and representation all are required to enable effective local interventions based on the open or other data.</p>
<p>7. Governance — the required financing, legal, regulatory or policy regime, required to enable the use to which the data would be put.</p>

Hoewel een dergelijke invalshoek binnen dit onderzoek niet aan bod komt en Gursteins model niet in het verdere onderzoek is meegenomen, is het nuttig om kritische tegengeluiden op te nemen in het theoretisch debat, om uiteindelijk tot beschikbare en toegankelijke open data te komen die daadwerkelijk universeel toegankelijk is.

3. Instelling & Aanbieder

3.1 Inleiding

Het OpenGLAM-onderzoek uit 2014-2016⁴ richt zich op digitale innovaties in de erfgoedsector zoals digitalisering van de collectie, open data en crowdsourcing. Dit onderzoek maakt gebruik van de innovatietheorie van Everett M. Rogers⁵ waarin verschillende stadia worden onderscheiden die een indicatie kunnen geven van veranderende trends binnen een vakgebied. Het onderzoek is uitgezet onder een groot aantal erfgoedinstellingen in Polen, Finland en Zwitserland en Nederland.

“...open data and open content are just about to diffuse; with adoption rates of 26% and 17% respectively. ...open data and open content are expected to diffuse relatively rapidly as these practices are considered as important by almost half of the respondents” (Estermann 1).

De ontwikkelingen op het gebied van open data en open content lijken binnen de erfgoedsector steeds meer geaccepteerd en langzamerhand steeds breder doorgevoerd te worden. Estermann et al. onderzochten verschillende verbanden tussen de adoptie van open data enerzijds en verschillende karakteristieken van de betreffende erfgoedinstelling anderzijds. Dit is gedaan aan de hand van de resultaten van een onderzoeksenquête. Van invloed op de verspreiding van digitale innovaties bij erfgoedinstellingen bleken onder andere: land, type instelling, type materiaal, geografisch bereik, grootte instelling, vaardigheden van het personeel en de houding van de instelling ten opzichte van beschikbaarstelling van content voor hergebruik (1, 11). Omdat Nederland volgens het onderzoek sterk voorop loopt in de adoptie van open data en open content (10), is het interessant om de resultaten uit het OpenGLAM-onderzoek voor Nederland opnieuw en diepgaander te analyseren. Het doel, van de secundaire analyse die is uitgevoerd, is na te gaan wat de karakteristieken zijn van Nederlandse erfgoedinstellingen die voorlopen bij de publicatie van open metadata en open content.

Hoewel in het OpenGLAM-onderzoek ook gegevens zijn verzameld over digitalisering, social media, crowdsourcing en linked data, worden deze gezien het kader van de huidige onderzoeksvraag niet meegenomen. Voor dit onderzoek definiëren we een drietal kernbegrippen, afgeleid van de Nederlandstalige OpenGLAM-survey:

Open metadata zijn op het internet beschikbaar gestelde digitale gegevens die worden gebruikt om de erfgoedobjecten in het bezit van een instelling te beschrijven die vrij mogen worden gebruikt, aangepast en gedeeld door iedereen voor elk doel, inclusief commercieel gebruik.

Open content zijn op het internet beschikbaar gestelde digitale kopieën / beelden van erfgoedobjecten die vrij mogen worden gebruikt, aangepast en gedeeld door iedereen voor elk doel, inclusief commercieel gebruik.

Open data zijn op het internet beschikbaar gestelde gegevens in een formaat dat leesbaar is voor computers en door iedereen voor elk doel mogen worden gebruikt. Met andere woorden: open data bestaat uit zowel open metadata als open content.

⁴ Zie: https://outreach.wikimedia.org/wiki/GLAM/OpenGLAM_Benchmark_Survey

⁵ *Diffusion of Innovations*, Everett M. Rogers, 1962.

3.2 Analyse en verrijking van de onderzoeksdata

Er zijn in Nederland 1393 instellingen benaderd voor het OpenGLAM-onderzoek. De algemene responsgraad was 12,3%; voor archieven was dit 16,3%, voor bibliotheken 14,6% en voor musea 9,8%. Het Nederland-specifieke databestand van het OpenGLAM-onderzoek bevat alle antwoorden op de vragen die zijn gesteld in de enquête⁶. Dit databestand bevat de responsdata van 292 Nederlandse instellingen. Hieruit zijn de erfgoedinstellingen geselecteerd die van de vragenlijst tenminste 20 van de 24 onvoorwaardelijke vragen hebben ingevuld. Tabel 1 weergeeft de 146 resterende instellingen ingedeeld naar instellingstype.

Instellingstype	Aantal
Archief	48 (33%)
Museum	67 (46%)
Bibliotheek	5 (3%)
Combinatie/Overig	26 (18%)
Totaal	146 (100%)

Tabel 1. Geselecteerde instellingen in OpenGLAM-onderzoek ingedeeld naar instellingstype.

In de OpenGLAM-gegevens blijken de Nederlandse bibliotheken sterk ondervertegenwoordigd; slechts 5 bibliotheken hebben de enquête volledig ingevuld. Daarnaast zijn er 26 instellingen (18%) die niet onder één instellingstype kunnen worden ondergebracht; daardoor zijn zij moeilijk met andere instellingen en sectoren te vergelijken. Archiven en musea maken samen 79% uit van de steekproef (115 erfgoedinstellingen); voor dit onderzoek is besloten om de verdere analyse tot deze twee instellingstypen te beperken.

De erfgoedinstellingen zijn vervolgens ingedeeld in drie categorieën van grootte (klein, middel, groot) om te kunnen analyseren of de grootte van de instelling verband houdt met adoptie van open data. De grootte van de erfgoedinstelling is in dit onderzoek afgeleid uit de antwoorden van het OpenGLAM-onderzoek op de vragen over omzet (vraag A8) en het personeel (vraag A6) waarbij alleen is gekeken naar betaalde fte's; vrijwilligers zijn niet meegeteld. Ook het geografische bereik (vraag A5) is in deze operationalisering niet meegenomen.

De inschaling voor het personeel in fte is als volgt bepaald:

- Klein: Tot 10 fte;
- Middel: Tussen 11 en 100 fte;
- Groot: Meer dan 100 fte.

De inschaling voor de omzet is als volgt bepaald:

- Klein: Tot €50.000;
- Middel: Tussen €50.000 en €500.000;
- Groot: Meer dan €500.000.

Deze twee variabelen zijn tegenover elkaar uitgezet om een inschaling te bepalen voor de grootte van de erfgoedinstellingen (tabel 2). Indien voor een instelling geen omzet uit de onderzoeksgegevens bekend was, prevaleerde het aantal fte's voor de bepaling van de instellingsgrootte.

⁶ Zie: http://survey.openglam.ch/data/The_Netherlands/GLAM2014_The_Netherlands_20150611_data_v0-94.sav

	Fte klein	Fte middel	Fte groot
Omzet klein	klein	klein	middel
Omzet middel	middel	middel	groot
Omzet groot	middel	groot	groot

Tabel 2. Inschaling van erfgoedinstellingen in grootte op basis van omzet en fte.

Deze inschaling is toegepast op het OpenGLAM-databestand en levert voor de erfgoedsectoren de volgende verdeling in grootte op:

Sector	Klein	Middel	Groot	Totaal
Archief	26 (54,2%)	14 (29,2%)	8 (16,7%)	48 (100%)
Museum	32 (47,8%)	21 (31,3%)	14 (20,9%)	67 (100%)
Totaal	58 (50,4%)	35 (30,4%)	22 (19,1%)	115 (100,0%)

Tabel 3. Nederlandse erfgoedinstellingen ingedeeld naar sector.

3.3 Analyse van adoptie van open data

De eerste statistische analyse in dit onderzoek is uitgevoerd om te onderzoeken of er sprake is van een verband tussen de adoptie van open data en de grootte van de instelling. Om adoptie te kunnen gebruiken in deze analyse is het van belang om deze eerst te definiëren en meetbaar te maken. Dat is in het OpenGLAM-onderzoek gedaan door het ontwikkelingsstadium ‘adoptie’ in de cyclus van stadia te definiëren aan de hand van antwoorden op specifieke vragen uit het OpenGLAM-onderzoek (Estermann 7-8). In dit ontwikkelingsstadium ‘adoptie’ volgens Estermann is ook de prognose die de instellingen geven over vijf jaar t.a.v. open data meegenomen. Om adoptie verder schaalbaar te maken in een adoptiegraad is voor dit onderzoek in eerste instantie een driedeling gemaakt met de volgende afleidingsregels met als achterliggende gedachte de voorlopers in beeld te kunnen brengen (de instellingen die het verst gevorderd zijn met open data):

Adoptiegraad 0: Beperkte adoptie

Beperkt aanbod open data, lage verwachtingen groei open data-aanbod van eigen instelling in de komende 5 jaar. Anders gezegd krijgt een instelling adoptiegraad 0 toegewezen indien niet aan de voorwaarden voor adoptiegraad 1 of 2 wordt voldaan.

Adoptiegraad 1: Enige adoptie

Open metadata aanwezig (tenminste 10% van beschikbare metadata opengesteld) met een groeiverwachting van tenminste 10% in de komende vijf jaar. Open content aanwezig (tenminste 10% van beschikbare erfgoedobjecten opengesteld) met een groeiverwachting van tenminste 5% in de komende vijf jaar.

Adoptiegraad 2: Vergaande adoptie (Voorlopers)

Tenminste 50% van de metadata en 50% van de beschikbare erfgoedobjecten zijn als open data beschikbaar gemaakt.

De adoptiegraad is voor alle deelnemende Nederlandse erfgoedinstellingen op de hiervoor beschreven manier afgeleid en vervolgens afgezet tegen de erfgoedsectoren en hun grootte (tabel 4).

Instellingstype	Grootte	Beperkte adoptie	Enige adoptie	Vergaande adoptie
Archief	klein n = 26	21	4	1
	middel n = 14	9	5	0
	groot n = 8	5	2	1
Musea	klein n = 32	23	7	2
	middel n = 21	12	7	2
	groot n = 14	6	5	3

Tabel 4. Open data adoptie bij Nederlandse erfgoedinstellingen ingedeeld naar sector en grootte.

Uit deze tabel kan voorzichtig worden afgeleid dat Nederlandse musea verder gevorderd lijken met het publiceren van open data dan archieven. Beide sectoren zijn in elk geval wel met open data bezig. Tussen kleine en middelgrote erfgoedinstellingen is echter weinig verschil te zien. Omdat de aantallen met adoptiegraad 2 zeer gering zijn, lijkt de aanvankelijke opzet voor het onderscheid in adoptiegraad 1 en 2 weinig zinvol.

Het lijkt, gezien de onderzoeksvraag, met name belangrijk onderscheid te maken tussen archieven en musea enerzijds en tussen open metadata en open content anderzijds. De operationalisering van adoptie volgens Estermann is daarom ingeruild voor een definitie die directer aansluit op de actuele adoptiestatus van een instelling. De rol van de door de instelling verwachte groei in het eigen open data-aanbod is daarbij komen te vervallen. In plaats daarvan wordt dus voor de adoptie voor de verdere OpenGLAM-analyse uitsluitend gekeken naar het huidige aanbod van open data.

De percentages in tabel 5 zijn berekend door per erfgoedinstelling de opgegeven percentages van open metadata (vraag C1, drie categorieën) en open content (vraag D4, acht categorieën) te middelen en deze twee gemiddelden aan het databestand toe te voegen als ‘adoptiepercentage open metadata’ en ‘adoptiepercentage open content’. Bij deze middeling zijn de open data-categorieën onderling als gelijkwaardig beschouwd, omdat in de enquête niet is gevraagd naar concrete aantallen erfgoedobjecten die een zinvolle waardering van categorieën mogelijk zou kunnen maken. Bijvoorbeeld archiefbronnen zouden bij archieven zwaarder gewogen kunnen worden dan geografische bronnen.

Bij de middeling van de percentages open metadata is rekening gehouden met de soorten metadata die de instelling in de enquête naar eigen zeggen bezit. Bij open content gaf de enquête geen optie aan of de instelling het type open content bezit; daarom is in dit onderzoek bij open content aangenomen dat als een instelling een percentage heeft opgegeven (ook al is dit nul procent) zij daarmee impliciet aangeven de categorie in huis te hebben en dat dit percentage daarmee in de middeling wordt meegewogen. Per sector (archieven versus musea) zijn uiteindelijk de gemiddelde percentages open metadata en open content wederom gemiddeld en naast elkaar gezet (zie tabel 5).

	Instellingstype					
	Archieven			Musea		
Open data-vorm	n	M	SD	n	M	SD
Open metadata	36	47,0%	37,6	43	38,7 %	38,3
Open content	43	16,3%	22,6	58	21,1 %	31,5

Tabel 5. Gemiddelde adoptiepercentages open metadata en open content per instellingstype.

Er is een t-test uitgevoerd met onafhankelijke steekproeven om vast te stellen of er verschillen waren tussen archieven en musea met betrekking tot de adoptiepercentages van open metadata en van open content. Er is gezocht naar outliers aan de hand van een boxplot; de vier gevonden outliers zijn voor deze t-test verwijderd. Hoewel archieven bij de gemiddelde adoptiepercentages van open metadata gemiddeld 8,3% hoger scoren (95% CI [-8,7 tot 25,4]), blijkt dit verschil niet statistisch significant, $t(77) = ,971$, $p = ,334$. Er blijkt wel een significant verschil in gemiddelde adoptiepercentages van open content tussen archieven en musea, waar archieven lager scoorden dan musea, $M = -11,0\%$, 95% CI [-20,2 tot -1,8], $t(81,082) = -2,389$, $p = ,019$.

Omdat onze onderzoeksvraag zich specifiek richt op de voorlopers wat betreft adoptie van open data hebben we voor de verdere analyse per instellingstype (archieven/musea) en open data-soort (metadata/content) gekeken naar de instellingen die boven het derde kwartiel scoorden op het percentage opengesteld erfgoed. Dit levert de onderstaande tabel 6 op.

	Voorloper	Adoptie open metadata	Adoptie open content
Archieven	Ja	$\geq 80\%$ n = 10	$\geq 25\%$ n = 11
	Nee	$\geq 80\%$ n = 38	$< 25\%$ n = 37
Musea	Ja	$\geq 80\%$ n = 12	$\geq 40\%$ n = 20
	Nee	$< 80\%$ n = 55	$< 40\%$ n = 47

Tabel 6. Percentages waarboven de adoptie van open data is aangemerkt als 'voorloper' per sector en type

Er zijn twee extra dichotome variabelen gemaakt (geen/wel voorloper) en aan het databestand (zie bijlage I) toegevoegd om te kunnen kijken naar verschillen tussen de groepen voorlopers en niet-voorlopers voor specifieke kenmerken van erfgoedinstellingen. Omdat niet alle instellingen de gebruikte vragen hebben beantwoord is het aantal instellingen dat uiteindelijk is meegenomen in de onderstaande berekeningen afhankelijk van zowel het instellingstype als het open data-type; deze aantallen zijn daarom ook vermeld in tabel 6.

Op de volgende pagina wordt per instellingstype en per open-datatype de top-10 van erfgoedinstellingen vermeld die volgens voorgaande afleiding voorloper zouden zijn in de adoptie van open data. Het is goed te beseffen dat deze volgorde sterk afhankelijk is van hoe de instellingen zelf hebben ingeschat welk percentage van hun collectie als open data beschikbaar is. Ook kan een instelling met een kleine collectie snel tot een hoge score komen. Omdat er geen kwantitatieve gegevens bekend zijn van de grootte van de collectie van een erfgoedinstelling kan dit helaas niet meewegen in de bepaling van het adoptiepercentage.

Voorlopers top-10 archieven open metadata (OMD)		
Archiefinstelling	OMD adoptie %	Grootte instelling
Historisch Centrum Leeuwarden	100	groot
Regionaal archief Alkmaar	100	groot
Data Archiving and Networked Services (DANS)	100	groot
Brabant-Collectie te Tilburg	100	klein
Nationaal Archief	98,33	groot
Oud Limburgse Schuttersfederatie	95	klein
Gemeentearchief Tholen	95	klein
Gemeentearchief Roosendaal	90	middel
Stadsarchief Mechelen	80	klein
Molentichting Goeree-Overflakkee	80	klein
Voorlopers top-10 archieven open content (OC)		
Archiefinstelling	OC adoptie %	Grootte instelling
Demijnen	100	klein
Indie Gangers	75	klein
Nationaal Archief	65	groot
Stichting Militair Erfgoed Groot-Amsterdam (MEGA)	50	klein
Collectie Brands	48	klein
Oud Limburgse Schuttersfederatie	37,5	klein
GeCa Archiefdienstverlening te Enschede	33	middel
Gemeentearchief Veldhoven	30	middel
Stichting Historie Ruinerwold	30	klein
Stichting Industrieel Erfgoed Deventer (SIED)	27	klein
Voorlopers top-10 musea open metadata (OMD)		
Archiefinstelling	OMD adoptie %	Grootte instelling
Museum Bredius te Den Haag	100	klein
Theepottenmuseum Swartbroek	100	klein
Onderzeebootloods te Rotterdam	98	klein
Nederlands Volksbuurtmuseum te Utrecht	96	klein
Naturalis te Leiden	95	groot
Flipje & Streekmuseum te Tiel	95	middel
Stedelijk Molenmuseum De Valk te Leiden	95	klein
Museum Flehite Amersfoort	94	groot
Museum Rotterdam	85	groot
Centraal Museum - Rietveld Schröderhuis te Utrecht	80	groot
Voorlopers top-10 musea open content (OC)		
Archiefinstelling	OC adoptie %	Grootte instelling
Onderzeebootloods te Rotterdam	98	klein
Flipje & Streekmuseum te Tiel	96,5	middel
Museum Flehite Amersfoort	94	groot
Streekmuseum Schippersbeurs te Elsloo	90	klein
Het Vreeswijks museum	84	klein
Industrieel Smalpoort Museum / Stichting Industrieel	80	klein
Tabaks Historisch Museum Delft	80	klein
Nederlands Volksbuurtmuseum te Utrecht	79	klein
Museum Catharijneconvent te Utrecht	70	klein
Museum Lichtschip 12 Noord Hinder te Hellevoetsluis	50	klein

3.4 Hypothesen, statistische analyse en bevindingen

Een aantal hypothesen zijn geformuleerd over mogelijke verbanden tussen adoptie en andere variabelen in het door ons verrijkte OpenGLAM-databestand. De hypothesen zijn aan de hand van een geformuleerde vraag statistisch getoetst in SPSS. Bij de Mann-Whitney U-test is naar aanleiding van Fritz, Morris en Richler (12) een effectgrootte berekend in de vorm van Cohens r : $r = z / \text{wortel van } N$. Daarbij is een groot effect ,50, een middelmatig effect ,30 en een klein effect ,10.

Hypothese 1: We vermoeden een verband tussen de grootte van de erfgoedinstelling (vraag A6, A8) en een hoge adoptie van open data. Dit wordt voor open metadata als positieve correlatie gesuggereerd in het OpenGLAM-onderzoek [Estermann, 10]. Uit hetzelfde onderzoek zou geen verband blijken voor open content.

Vraag 1: Is er een significant verschil tussen voorlopers en niet-voorlopers van open metadata / open content wat betreft de instellingsgrootte?

Uit een Mann-Whitney U-test blijkt bij open metadata een significant verschil met zwak tot middelmatig effect ($r = ,25$) tussen voorlopers (Med = 2 [middel]) en niet-voorlopers (Med = 1 [klein]) wat betreft de instellingsgrootte, $U = 1368$, $z = 2,682$, $p = ,007$. Als alleen de archieven worden geselecteerd blijft het verschil significant, $U = 284$, $z = 2,648$, $p = ,016$, en neemt de effectgrootte licht toe ($r = ,29$). Kijkend naar alleen musea vervalt de significantie, $U = 402$, $z = 1,276$, $p = ,202$; grote archieven zijn dus eerder voorlopers dan kleine archieven op het gebied van open metadata, terwijl bij musea geen verband is aangetoond.

De test is herhaald voor open content voor verschillen tussen voorlopers (Med = 1) en niet-voorlopers (Med = 2). Er blijken geen statistisch significante verschillen ($U = 1240$, $z = -,424$, $p = ,672$), ook niet wanneer wordt geselecteerd op musea ($U = 499$, $z = ,431$, $p = ,667$) of archieven ($U = 154$, $z = -1,361$, $p = ,173$).

Hypothese 2: We vermoeden een verband tussen een groot geografisch bereik en een hoge adoptie van open data.

Vraag 2: Is er een significant verschil tussen voorlopers en niet-voorlopers van open metadata / open content wat betreft geografisch bereik?

Uit een Mann-Whitney U-test blijkt wat betreft open metadata en geografisch bereik geen statistisch significant verschil tussen niet-voorlopers (Med = 2 [regionaal/nationaal]) en voorlopers (Med = 2) voor musea en archieven bij elkaar ($U = 1200$, $z = 1,341$, $p = ,180$) noch voor musea apart ($U = 410$, $z = 1,377$, $p = ,168$) noch voor archieven apart ($U = 219$, $z = ,821$, $p = ,475$). Ook wat betreft open content en geografisch bereik blijken geen statistisch significante verschillen ($U = 1252$, $z = -,336$, $p = ,737$).

Hypothese 3: We vermoeden een verband tussen het feit dat erfgoedinstellingen hebben aangegeven ‘tweedimensionale visuele bronnen’ te hebben, en een hoge adoptie van open content.

Vraag 3: Is er een significant verschil tussen voorlopers en niet-voorlopers van open content wat betreft het aangeven dat ‘tweedimensionale visuele bronnen’ kenmerkend zijn voor hun instelling?

Uit een chi-kwadraattoets blijken de verschillen tussen voorlopers en niet-voorlopers noch voor archieven ($\chi^2(1) = ,573$, $p = ,449$) noch voor musea ($\chi^2(1) = ,094$, $p = ,760$) noch voor archieven en musea bij elkaar ($\chi^2(1) = ,574$, $p = ,449$) significant.

Hypothese 4: We vermoeden een verband tussen het feit dat erfgoedinstellingen hebben aangegeven ‘bronnen gebaseerd op tekst’ te hebben, en een hoge adoptie van open metadata.

Vraag 4: Is er een significant verschil tussen voorlopers en niet-voorlopers van open metadata wat betreft het aangeven dat ‘bronnen gebaseerd op tekst’ kenmerkend zijn voor hun instelling?

Uit een chi-kwadraattoets blijken de verschillen tussen voorlopers en niet-voorlopers noch voor archieven ($\chi^2(1) = 1,049$, $p = ,306$) noch voor musea ($\chi^2(1) = -,123$, $p = ,316$) noch voor archieven en musea bij elkaar ($\chi^2(1) = 1,681$, $p = ,195$) significant.

Hypothese 5: We vermoeden een verband tussen het feit dat erfgoedinstellingen aangeven dat de belangrijkste gebruikers ‘Onderzoeksinstellingen/specialisten’ zijn en een hoge adoptie van open metadata.

Vraag 5: Is er een significant verschil tussen voorlopers en niet-voorlopers van open metadata wat betreft het aangeven dat de belangrijkste gebruikers ‘Onderzoeksinstellingen/specialisten’ zijn?

Uit een chi-kwadraattoets blijken de verschillen tussen voorlopers en niet-voorlopers noch voor archieven ($\chi^2(1) = 2,468$, $p = ,116$) noch voor musea ($\chi^2(1) = ,010$, $p = ,920$) noch voor archieven en musea bij elkaar ($\chi^2(1) = 1,617$, $p = ,204$) significant.

Hypothese 6: We vermoeden een verband tussen het feit dat erfgoedinstellingen aangeven van publieke fondsen afhankelijk te zijn en een hoge adoptie van open metadata en open content.

Vraag 6: Is er een significant verschil tussen voorlopers en niet-voorlopers van open metadata / open content wat betreft de afhankelijkheid van publieke fondsen te zijn?

Uit een t-test met onafhankelijke steekproeven blijkt bij open metadata een statistisch significant verschil tussen voorlopers ($M = 60,1$, $SD = 39,9$) en niet-voorlopers ($M = 28,3$, $SD = 37,6$) wat betreft afhankelijkheid van publieke fondsen, $M = -31,8$, $SE = 9,0$, $t(111) = -3,515$, $p = ,001$. Dat verschil blijkt ook significant bij selectie van alleen musea ($M = -24,9$, $SE = 10,7$, $t(63) = -2,330$, $p = ,023$) of alleen archieven ($M = -38,9$, $SE = 15,0$, $t(46) = -2,602$), $p = ,012$).

De t-test is herhaald voor open content. Er blijkt hier geen significant verschil tussen voorlopers ($M = 37,1$, $SD = 41,0$) en niet-voorlopers ($M = 27,7$, $SD = 36,8$): $M = 9,4$, $SE = 8,4$, $t(111) = 1,122$, $p = ,264$. Bij selectie van alleen archieven is er wel een significant verschil ($M = 31,8$, $SE = 14,7$, $t(46) = 2,160$, $p = ,036$). Bij selectie van alleen musea is er geen significant verschil ($M = -7,0$, $SE = 9,3$, $t(63) = -,752$, $p = ,455$).

Hypothese 7: We vermoeden een verband tussen het feit dat erfgoedinstellingen hebben aangegeven van verkoop van beeldrechten afhankelijk te zijn en een hoge adoptie van open metadata en open content.

Vraag 7: Is er een significant verschil tussen voorlopers en niet-voorlopers van open metadata / open content wat betreft het aangeven van verkoop van beeldrechten afhankelijk te zijn?

Uit een t-test met onafhankelijke steekproeven blijkt bij open metadata geen significant verschil tussen voorlopers ($M = ,23$, $SD = 1,26$) en niet-voorlopers ($M = ,18$, $SD = ,501$): $t(113) = ,160$, $M = ,04$, $SE = ,27$, $p = ,795$. Er is ook geen significant verschil bij selectie van musea ($t(65) = -,750$, $p = ,456$) of van archieven ($t(46) = ,559$, $p = ,579$). Bij open content zijn eveneens geen significante verschillen aangetroffen ($t(113) = ,134$, $p = ,894$).

Hypothese 8: We vermoeden een verband tussen de organisatorische vorm van de instelling en een hoge adoptie van open metadata en open content.

8) Is er een significant verschil tussen voorlopers en niet-voorlopers van open metadata wat betreft de organisatorische vorm van de instelling?

Omdat er slechts 2 instellingen zich een commerciële instelling noemden en 7 instellingen een gemengde vorm hebben, zijn deze niet meegerekend bij het beantwoorden van deze vraag. Er is dus alleen gekeken naar het verschil tussen publieke en private instellingen.

Uit een chi-kwadraattoets volgt een significant zwak verband tussen voorlopers en niet-voorlopers op het gebied van open metadata, kijkend naar musea en archieven bij elkaar genomen ($\chi^2(1) = 3,868$, $\phi = ,191$, $p = ,049$). Dit verband wordt sterker als archieven apart worden genomen ($\chi^2(1) = 4,220$, $\phi = ,306$, $p = ,040$) en blijkt niet meer significant als musea apart worden genomen ($\chi^2(1) = ,282$, $\phi = ,068$, $p = ,595$). Publieke archieven blijken dus vaker voorlopers op het gebied van open metadata dan private instellingen. Er blijkt wat betreft open content geen significant verband tussen voorlopers en niet-voorlopers voor archieven ($\chi^2(1) = 2,701$, $\phi = ,245$, $p = ,100$), noch voor musea ($\chi^2(1) = ,753$, $\phi = ,111$, $p = ,386$), noch voor musea en archieven bij elkaar ($\chi^2(1) = ,302$, $\phi = ,053$, $p = ,583$).

Hypothese 9: We vermoeden een verband tussen het feit dat er goedgeinstellingen hebben aangegeven 'bronnen gebaseerd op tekst' te hebben, en een hoge adoptie van open metadata.

9) Is er een significant verschil tussen voorlopers en niet-voorlopers van open metadata wat betreft het aangeven dat 'bronnen gebaseerd op tekst' kenmerkend zijn voor hun instelling?

Uit een chi-kwadraattoets blijken de verschillen tussen voorlopers en niet-voorlopers niet significant ($\chi^2(1) = 1,681$, $\phi = ,121$, $p = ,195$) met archieven en musea tezamen; ook kijkend naar alleen musea ($\chi^2(1) = 1,006$, $\phi = ,123$, $p = ,316$) of alleen archieven ($\chi^2(1) = 1,049$, $\phi = ,148$, $p = ,306$) blijken geen statistisch significante verschillen.

Hypothese 10: We vermoeden een verband tussen het feit dat een instelling aangeeft dat het veel belang hecht aan open metadata en een hoge adoptie van open metadata.

Vraag 10: Is er een significant verschil tussen voorlopers en niet-voorlopers van open metadata wat betreft het aangeven door de instelling dat het veel belang hecht aan open metadata?

Uit een Mann-Whitney U-test blijkt een significant verschil met middelmatig effect ($r = ,33$) tussen voorlopers (Med = 4,5) en niet-voorlopers (Med = 4), $U = 1102$, $z = 3,200$, $p = ,001$. Bij een selectie van alleen musea neemt de effectgrootte licht toe ($r = ,36$), $U = 344,5$, $z = 2,704$, $p = ,007$. Bij een selectie van alleen archieven blijkt geen significant verschil ($U = 209$, $z = 1,724$, $p = ,105$).

Hypothese 11: We vermoeden een verband tussen het feit dat een instelling aangeeft dat het veel belang hecht aan open content en een hoge adoptie van open content.

Vraag 11: Is er een significant verschil tussen voorlopers en niet-voorlopers van open content wat betreft het aangeven door de instelling dat het veel belang hecht aan open content?

Uit een Mann-Whitney U-test blijkt een significant verschil met middelmatig effect ($r = ,37$) tussen voorlopers (Med = 5) en niet-voorlopers (Med = 3), $U = 1453$, $z = 3,682$, $p < ,0005$. Bij een selectie van alleen musea neemt de effectgrootte licht af ($r = ,34$), $U = 492$, $z = 2,521$, $p = ,012$. Bij een selectie van alleen archieven neemt het effect sterk toe ($r = ,46$), $U = 250,5$, $z = 3,030$, $p = ,002$.

Hypothese 12: We vermoeden een verband tussen het feit dat een instelling aangeeft dat open metadata kansrijk is en een hoge adoptie van open metadata.

Vraag 12: Is er een significant verschil tussen voorlopers en niet-voorlopers van open metadata wat betreft het aangeven door de instelling dat open metadata kansrijk is?

Uit een Mann-Whitney U-test blijkt geen significant verschil tussen voorlopers (Med = 4) en niet-voorlopers (Med = 4), $U = 879,5$, $z = ,912$, $p = ,362$. Ook bij selectie van alleen musea ($U = 256$, $z = ,713$, $p = ,476$) of alleen archieven ($U = 176$, $z = ,332$, $p = ,740$) blijkt geen significant verschil.

Hypothese 13: We vermoeden een verband tussen het feit dat een instelling aangeeft dat open content kansrijk is en een hoge adoptie van open content.

Vraag 13: Is er een significant verschil tussen voorlopers en niet-voorlopers van open content wat betreft het aangeven door de instelling dat open content kansrijk is?

Uit een Mann-Whitney U-test blijkt een significant verschil met zwak effect ($r = 0,21$) tussen voorlopers (Med = 4) en niet-voorlopers (Med = 4), $U = 1216$, $z = 2,007$, $p = ,045$. Bij selectie van alleen

musea ($U = 375,5$, $z = 1,014$, $p = ,311$) noch bij selectie van alleen archieven ($U = 236,5$, $z = 1,991$, $p = ,046$) blijkt geen significant verschil.

Hypothese 14: We vermoeden een verband tussen het feit dat een instelling aangeeft dat medewerkers professioneel betrokken zijn bij content creatie voor Wikipedia (en/of Wikimedia Commons / Wikidata) en een hoge adoptie van open content.

Vraag 14: Is er een significant verschil tussen voorlopers en niet-voorlopers van open content wat betreft het aangeven door de instelling dat medewerkers betrokken zijn bij content creatie op Wikipedia en/of Wikimedia Commons en/of Wikidata?

Uit een chi-kwadraattoets blijken de verschillen tussen voorlopers en niet-voorlopers niet significant ($\chi^2(1) = 2,662$, $\phi = ,152$, $p = ,103$) met archieven en musea tezamen; ook kijkend naar alleen musea ($\chi^2(1) = 1,530$, $\phi = ,151$, $p = ,216$) of alleen archieven ($\chi^2(1) = ,922$, $\phi = ,139$, $p = ,337$) blijken geen statistisch significante verschillen.

Andere vragen en hypothesen die we vanwege de beperkte onderzoekstijd niet hebben kunnen beantwoorden:

- Is er een verband tussen het feit dat een instelling aangeeft dat er veel gedigitaliseerd is en een hoge adoptie van open data? (vraag D-1)
- Is er een verband tussen het feit dat een instelling aangeeft delen van de collectie niet te digitaliseren om reden van gebrek aan financiering en een hoge adoptie van open data? (vraag D-2)
- Is er een verband tussen het feit dat een instelling aangeeft dat content gratis beschikbaar gemaakt moet worden voor onderwijs en onderzoek en een hoge adoptie van open data? (vraag D-3)
- Is er een verband tussen het feit dat een instelling aangeeft dat content gratis beschikbaar gemaakt moet worden voor commerciële gebruikers en een hoge adoptie van open data? (vraag D-3)
- Is er een verband tussen het feit dat een instelling aangeeft veel content te hebben in het publieke domein en een hoge adoptie van open data? (vraag D-5)
- Is er een verband tussen het feit dat een instelling aangeeft absoluut voordeel te zien in de mogelijkheid om de zichtbaarheid en/of relevantie van de instelling te verbeteren en een hoge adoptie van open content? (vraag D-7)
- Is er een verband tussen het feit dat een instelling aangeeft weinig uitdaging te hebben in tijd en kosten gepaard gaande met digitalisering en een hoge adoptie van open content? (vraag D-8)
- Is er een verband tussen het feit dat een instelling aangeeft weinig risico te zien tav verkeerd gebruik of representatie van de content en een hoge adoptie van open content? (vraag D-9)
- Is er een verband tussen het feit dat een instelling aangeeft publiek te bereiken via social media op collaborative online communities zoals Wikipedia en een hoge adoptie van open metadata en open content? (vraag E-1)
- Is er een verband tussen voldoende aanwezige vaardigheden van het personeel en een hoge adoptie van open data en open content? (Vraag G-1)

3.5 Korte samenvatting en conclusies

Er is een nadere analyse uitgevoerd naar de adoptie van open data bij Nederlandse erfgoedinstellingen. Dit naar aanleiding van het artikel van Estermann en de beschikbare antwoordgegevens op vragen uit een enquête gehouden onder erfgoedinstellingen. Bij deze secundaire analyse is de centrale vraag geworden wat de karakteristieken zijn van erfgoedinstellingen in de sectoren archieven en musea die voorlopen in de adoptie van open data.

Om deze analyse te kunnen uitvoeren zijn definities van open data, van de grootte van de erfgoedinstellingen en van adoptie van open data opnieuw vastgesteld. Open data bestaat uit open metadata

en open content; de grootte kan worden afgeleid uit fte en omzet van de instelling; de adoptie kan worden afgeleid uit de percentages open metadata en open content die door de instellingen zijn ingevuld in de OpenGLAM-enquête.

De volgende gegevens zijn tijdens de analyse per instelling afgeleid en aan het databestand toegevoegd: grootte van de instelling (klein/middel/groot), percentage adoptie open metadata, percentage adoptie open content, indicator voorloper open metadata, indicator voorloper open content. Uit een t-test waarin de voorlopers van open data zijn vergeleken met de rest van de instellingen blijkt een significant verschil in de adoptie van open data tussen de instellingstypen. De musea laten hierbij een hogere adoptie zien voor open content dan de archieven. Voor open metadata is er geen verschil te zien tussen de sectoren. Een groot aantal gedetailleerde onderzoeksvragen met betrekking tot de data is opgesteld, waarvan een deel via statistische analyse is onderzocht. Omwille van de beperkte tijd zijn niet alle vragen aan de orde gekomen. Dit is voor vervolgonderzoek zeker interessant.

Conclusies over verbanden tussen adoptie van open data en instellingskarakteristieken

De volgende significante verbanden zijn gevonden:

- (1) Grotere archief-instellingen zijn eerder voorlopers dan kleinere archief-instellingen op het gebied van open metadata; voor musea is geen significant verband gevonden. Wat betreft open content is er noch voor archieven noch voor musea een significant verband gevonden.
- (6) Als instellingen (archieven of musea) aangeven van publieke fondsen afhankelijk te zijn, zijn zij eerder voorloper van open metadata; Bij archieven blijkt er ook dit verband te zijn bij open content, maar dat is bij musea niet het geval.
- (8) Publieke archiefinstellingen zijn eerder voorloper van open metadata dan private instellingen. Bij musea is geen verband gevonden met de organisatievorm van de instelling.
- (10) Musea die veel belang hechten aan open metadata zijn eerder voorloper van open metadata dan musea die weinig belang hechten; dat verband is bij archieven niet gevonden.
- (11) Archieven die veel belang hechten aan open content zijn eerder voorloper van open content dan archieven die weinig belang hechten; dat verband is bij musea niet gevonden.
- (13) Er is een zwak verband gevonden tussen de voorlopers van open content en het feit of instellingen hebben aangegeven dat zij open content kansrijk vinden, maar als we dat per instellingstype bekijken (archieven of musea) dan blijkt dat verband eigenlijk niet. We kunnen niet met zekerheid zeggen dat instellingen die open content kansrijk vinden ook voorloper van open content zijn.

Er zijn geen significante verschillen gevonden tussen voorlopers en niet-voorlopers:

- (2) bij open data en het geografisch bereik van erfgoedinstellingen;
- (3) bij open content en het aangeven van 'tweedimensionale visuele bronnen' als kenmerkend voor de instelling;
- (4) bij open metadata en het aangeven van 'bronnen gebaseerd op tekst' als kenmerkend voor de instelling;
- (5) bij open metadata en het aangeven van onderzoekinstellingen/specialisten als de belangrijkste gebruikers van die data;
- (7) bij open data en de mate van afhankelijkheid van inkomsten uit de verkoop van beeldrechten;
- (12) bij open metadata en de aangegeven mate van kansrijkheid van open metadata;
- (14) bij open content het aangeven van professionele betrokkenheid bij de content creatie op Wikipedia en/of Wikimedia Commons en/of Wikidata.

4. Gebruik: GLAMetrics

4.1. Inleiding

Erfgoedinstellingen zoals musea, bibliotheken en archieven hechten in toenemende mate waarde aan het gebruik en hergebruik van open content en open data in het publieke domein. Hierbij kan gedacht worden aan de websites van iedere instelling zelf maar ook websites die het delen van content mogelijk maken zoals Wikimedia Commons⁷ en Flickr Commons⁸. Het beschikbaar stellen in het publieke domein draagt erfgoedinstellingen bij in hun publieke taak ons gemeenschappelijke culturele erfgoed voor iedereen toegankelijk te maken - niet alleen via eigen kanalen maar ook daarbuiten. Daarnaast stimuleert het samenwerking tussen de erfgoedinstellingen wat nieuwe vormen van dienstverlening mogelijk maakt.

Een aantal erfgoedinstellingen die inmiddels open content en open data beschikbaar hebben gesteld, blijken naast het beschikbaar stellen ook geïnteresseerd in het vaststellen van de impact die dit beschikbaar stellen heeft. Hiermee wordt specifiek bedoeld op de mate van bereik en het hergebruik van de content. In het volgende hoofdstuk is een aantal erfgoedinstellingen geënquêteerd die hier evident blijken van te geven. Niet alleen is het meten “leuk om materiaal elders te zien opduiken”, zoals het Regionaal archief Nijmegen beweert⁹, het is ook een middel om vast te stellen of de open content en open data ook daadwerkelijk enig doel dient of gaat dienen. Het Gemeentearchief Rotterdam heeft als doel op een zo eenvoudig mogelijke manier een zo groot mogelijke groep gebruikers bij zoveel mogelijk open content te brengen. De instelling is hierbij van mening dat de verwachtingen rondom het “bereik en het gebruik van onze content vergroten” zijn uitgekomen¹⁰.

Voor wat betreft erfgoedinstellingen die delen van hun digitale collecties beschikbaar hebben gesteld op Wikimedia Commons, is in 2015 een verkennend onderzoek uitgevoerd door Open Cultuur Data (hierna: OCD)¹¹. Dit onderzoek richt zich op het aanbod van open content en open data voor het (her)gebruik op Wikipediaprojecten. Hierbij gaat het om digitale representaties van beeld-, video- en geluidsbestanden die via Wikimedia Commons vrij van auteursrecht aangeboden worden door middel van een Creative Commons-licentie. Het onderzoek heeft tot voorlopige bevindingen geleid die het resultaat zijn van een analyse van de webstatistieken over de periode van november 2014 tot en met mei 2015. Zo werd de verhouding van (digitale representaties van) het totale aanbod van Nederlandse erfgoedinstellingen op Wikimedia versus het totale aanbod geschat op 2,4%. De onderzoekers hebben aan de hand van enkele meetinstrumenten, die zij scharen onder de term “GLAMetrics” uitspraken gedaan over: 1) hergebruik, oftewel de mate waarin content daadwerkelijk een weg heeft gevonden van Wikimedia Commons naar Wikipedia en 2) bereik, oftewel de hoeveelheid raadplegingen van pagina's waarop hergebruikte content van de instellingen voorkomt.

4.2 Werkwijze en onderzoeksvraag

Voor ons onderzoek hebben wij gekozen voor het verrichten van een vervolgonderzoek op basis van voornoemd onderzoek van OCD uit 2015. Aangezien ons onderzoek zich richt op het huidige gecombineerde aanbod van content van Nederlandse erfgoedinstellingen, hebben wij gekozen voor de selectie van Nederlandse instellingen die content via Wikimedia Commons beschikbaar hebben gesteld. Hierbij hebben wij aanvullend gekozen voor een uitsplitsing naar type instelling om te traceren welke instellingen meer of minder gebaat zouden zijn bij het hergebruik van hun content op (onder andere)

⁷ Wikimedia Commons is een meertalige website die als repository dient voor beeld-, video- en geluidsbestanden. Zie: <https://commons.wikimedia.org>

⁸ Flickr Commons is een website die als repository dient voor het delen van beeld- en videobestanden. Zie: <https://www.flickr.com/commons>.

⁹ Verwijzing naar: enquête, vraag 4, p. 36

¹⁰ Verwijzing naar: enquête, vraag 4, p. 36

¹¹ Zie: <http://www.opencultuurdata.nl/2015/05/glametrics-nederlands-cultureel-erfgoed-op-wikipedia-bereikt-maandelijks-miljoenenpubliek>.

Wikipediaprojecten. Tevens is gekozen voor een opsplitsing van het soort content in beeld, video en geluid. Voorbeelden hierbij zijn digitale representaties van schilderijen, foto's, 3D-objecten, video en geluidsopnames.

Uiteindelijk moet ons onderzoek de onderstaande deelvraag beantwoorden:

Welke erfgoedinstellingen stellen op Wikimedia Commons open content en data beschikbaar voor (her)gebruik en hoe onderscheiden de voorlopers op dit terrein zich van de rest?

Uit deze deelvraag hebben we een aantal nadere vragen afgeleid die in ons onderzoek successievelijk zullen worden beantwoord:

- Wat zijn GLAMetrics en welke meetinstrumenten - tools - op dat terrein kunnen voor een analyse van webstatistieken van Wikipedia ingezet worden?
- Wat is met de beschikbare tools exact mogelijk en onmogelijk om te meten als het gaat om het hergebruik en bereik van open content en open data?
- Wat is precies de verhouding tussen musea, bibliotheken en archieven als het gaat om hun totale aanbod op Wikimedia Commons afgezet tegen het daadwerkelijk hergebruik op Wikipedia-projecten?
- Wat zijn de ontwikkelingen in één jaar tijd (van 31-12-2014 t/m 31-12-2015) als het gaat om aanbod en raadpleging. Aan welke instellingen is de grootste groei te relateren en zijn deze instellingen ook als “voorloper” te beschouwen?

4.3 Meetinstrumenten voor webstatistieken van Wikipedia

Er zijn verschillende meetinstrumenten - tools - beschikbaar om webstatistieken van Wikipedia te analyseren. Onder andere komen hiervoor in aanmerking: BaGLAMa 2¹², GLAMorous¹³ en Treeviews¹⁴.

Voor wat betreft de laatstgenoemde tool kan gesteld worden dat deze zich richt op de pageviews per categorieën (of groepen) van artikelen op Wikipedia. Deze tool lijkt een zekere verwantschap te hebben met de “article traffic statistics” tool, die via stats.grok.se/ wordt aangeboden. De statistische data van deze tools hebben geen betrekking op de via Wikimedia Commons aangeboden content, zodat ze buiten de scope van ons onderzoek vallen.

BaGLAMa 2 en GLAMorous richten zich beide specifiek op de content op Wikipedia via Wikimedia Commons. De eerstgenoemde tool richt zich op pageviews van pagina's waarop (hergebruikte) content uit Wikimedia Commons wordt gebruikt – de “Common files”. GLAMorous richt zich op het daadwerkelijk hergebruik afgezet tegen de totale hoeveelheid content dat per instelling wordt. Voor dat wij ingaan op de verzameling van meetgegevens van erfgoedinstellingen, zal hieronder de werking (toepassingsmogelijkheden) van iedere tool uiteengezet worden.

4.3.1 BaGLAMa 2

De tool BaGLAMa 2 presenteert de aangeboden objecten die aan ieder instelling gekoppeld kunnen worden. Een voorbeeld hierbij is “Collections of the Amsterdam Museum”¹⁵, een enkele categorie op Wikimedia Commons. Een korte blik op Wikimedia Commons leert ons dat van deze instelling content in zowel de (hoofd)categorie als subcategorieën is ondergebracht. BaGLAMa 2 neemt alle content in zowel de hoofd- als subcategorieën in ogenschouw. Met behulp van de tool kan vervolgens per tijdperiode van een maand het aantal pageviews van content van deze instelling, dat de verschillende Wikipediaprojecten wordt (her)gebruikt, worden bekeken. Onderstaande afbeelding geeft dit weer.

¹² Zie: <https://tools.wmflabs.org/glamtools/baglamma2>

¹³ Zie: <https://tools.wmflabs.org/glamtools/glamorous.php>

¹⁴ Zie: <https://tools.wmflabs.org/glamtools/treeviews/>

¹⁵ Zie: https://commons.wikimedia.org/wiki/Category:Collections_of_the_Amsterdam_Museum

Category details for Collections of the Amsterdam Museum

14 months have a data point, with 8,674,600 page views in total. Click on individual time points in the graph to see monthly data.



Page views in 2015-12

Total monthly page views: 805,316. [Download this table.](#)

Site	Pages	Views
English Wikipedia	Details	106 548,273
German Wikipedia	Details	54 56,225
Dutch Wikipedia	Details	146 51,682
Italian Wikipedia	Details	19 34,144

Afbeelding 1. Resultaten BaGLAMa 2 voor het Amsterdam Museum.

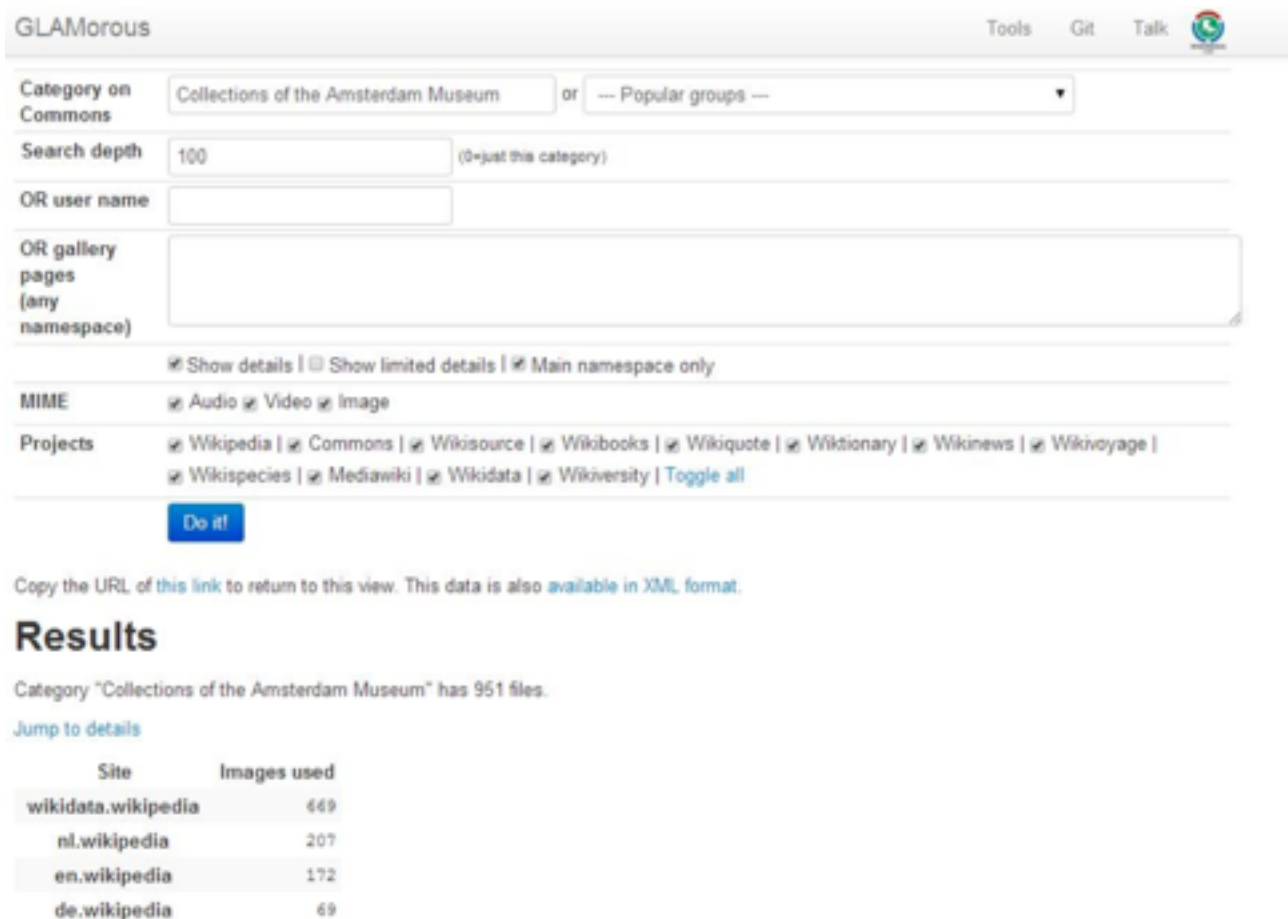
Hieruit valt in eerste instantie al op te maken dat content van het Amsterdam Museum voor de maand december 2015 verreweg het meest geraadpleegd wordt via de Engelstalige Wikipedia; Nederland staat hierbij op de derde plaats. Hieruit is af te leiden dat het museum met het aanbieden van content op Wikimedia Commons op dat moment een groot internationaal bereik heeft. De bijna 550.000 pageviews hebben betrekking op 106 pagina's waarop gebruik gemaakt wordt van content van het Amsterdam Museum op Wikimedia Commons. Via "details" kan nader inzichtelijk worden gemaakt op welke pagina's deze content terecht is gekomen en of er ook sprake is van één afbeelding op meerdere Wikipedia-pagina's. Gerangschikt per pagina wordt duidelijk welke afbeeldingen op welke pagina(s) het meest geraadpleegd worden. Uit de pageviews-totale blijkt dat pageviews weliswaar in verband kunnen worden gebracht met de beschikbare content van iedere instelling, maar dat de raadpleging in eerste instantie betrekking heeft op de het gebruik van een Wikipedia-pagina en niet de content zelf. Het doen van mogelijke uitspraken over de werkelijke reden van raadpleging door gebruikers van Wikipedia, is dus aan de hand van deze statistieken c.q. deze tool niet goed mogelijk. Er kan ook niet worden vastgesteld of gebruikers bij toeval een afbeelding zien, of dat zij gericht op zoek zijn naar de betreffende afbeelding. Ondanks deze beperkingen wordt door middel van BaGLAMa 2 wel inzichtelijk op welke wijze de door de diverse erfgoedinstellingen aangeboden content via Wikimedia Commons hun weg vindt naar de uiteindelijke gebruiker en of deze content mogelijk vaak of minder vaak op het scherm van de Wikipedia-gebruiker verschijnt.

4.3.2. GLAMorous

GLAMorous kan de beschikbare collecties van instellingen op Wikimedia Commons op een geheel andere wijze analyseren door hergebruik af te zetten tegen de totale hoeveelheid aangeboden content. De input voor

deze tool is iedere OCD-set die als “categorie” op Wikimedia te raadplegen is. Dit zijn dus ook de categorieën die onder andere in het overzicht van de BaGLAMa 2 worden weergegeven.

Voor de werking van deze tool is het van belang op te merken dat een volledige statistische analyse alleen mogelijk is wanneer de (hoofd)categorie inclusief alle subcategorieën meegenomen worden voor een analyse door de tool. Het veld “search depth” heeft hier betrekking op: het is de zoekdiepte van de subcategorieën. Een kort vooronderzoek liet ons zien dat van de door ons onderzochte Nederlandse erfgoedinstellingen die content via Wikimedia Commons aanbieden (zie hiertoe de volgende paragraaf) de zoekdiepte in alle gevallen niet meer dan 10 niveaus bedraagt. Met een ‘zekerheidsmarge’ van 100, moet naar onze mening alle content per instelling - voor analyse meegenomen worden. Deze marge hebben wij derhalve als standaardwaarde vastgesteld.



The screenshot shows the GLAMorous tool interface. At the top, there are navigation links for 'Tools', 'Git', 'Talk', and a logo. The main search area includes a dropdown menu for 'Category on Commons' set to 'Collections of the Amsterdam Museum' or 'Popular groups'. Below this is a 'Search depth' input field set to '100' with a note '(0=just this category)'. There are also fields for 'OR user name' and 'OR gallery pages (any namespace)'. Below these fields are checkboxes for 'Show details', 'Show limited details', and 'Main namespace only'. The 'MIME' section has checkboxes for 'Audio', 'Video', and 'Image'. The 'Projects' section lists various Wikimedia projects with checkboxes, including 'Wikipedia', 'Commons', 'Wikisource', 'Wikibooks', 'Wikiquote', 'Wiktionary', 'Wikinews', 'Wikivoyage', 'Wikispecies', 'MediaWiki', 'Wikidata', and 'Wikiversity', with a 'Toggle all' link. A blue 'Do it!' button is at the bottom of the search area. Below the search area, there is a link to 'Copy the URL of this link to return to this view. This data is also available in XML format.' The 'Results' section shows 'Category "Collections of the Amsterdam Museum" has 951 files.' and a 'Jump to details' link. A table with two columns, 'Site' and 'Images used', lists the following data:

Site	Images used
wikidata.wikipedia	669
nl.wikipedia	207
en.wikipedia	172
de.wikipedia	69

Afbeelding 2. Zoekscherm van GLAMorous voor de aangeboden content van het Amsterdam Museum.

De resultaten laten zien dat het Amsterdam Museum via Wikimedia in totaal 951 bestanden beschikbaar heeft gesteld voor (her)gebruik op Wikimedia. Van belang voor ons onderzoek is ook de waarde “Distinct images used” die onderaan de resultaten wordt weergegeven: in totaal worden hiervan 746 unieke bestanden hergebruikt. Op een totaal van 951 bestanden, die het museum (in één categorie op Wikimedia) heeft geplaatst, levert dat een percentage op van 78,44%. Wanneer dit percentage voor alle Nederlandse erfgoedinstellingen wordt berekend, zou het mogelijk moeten zijn uitspraken te doen over de mate waarin content daadwerkelijk wordt hergebruikt per instelling.

4.4 Onderzoeksmethode

Voor ons onderzoek maken we, gelet op het voorgaande, gebruik van BaGLAMa 2 in combinatie met GLAMorous om uiteindelijk het totale aanbod van open content en data per erfgoedinstelling af te zetten tegen het daadwerkelijke hergebruik op Wikipedia. Door dit vervolgens te vergelijken met de betreffende

pageviews per instelling in een bepaalde periode, wordt het mogelijk om uitspraken te doen over hergebruik en bereik en de ontwikkelingen in dat kader. Het samenspel van hergebruik en bereik (via pageviews) kan ons mogelijk antwoord geven op de vraag of er ook voorlopers aan te wijzen zijn.

Pageviews hebben wij kunnen opsplitsen in twee toetsmomenten: december 2014 en december 2015. Alhoewel wij ook dezelfde vergelijking voor het hergebruik in één jaar tijd in gedachten hadden, bleek dit via BaGLAMa 2 niet direct mogelijk te zijn. De tool presenteert namelijk alleen een actuele stand van zaken wat betreft het aanbod op Wikimedia Commons en het uiteindelijke hergebruik op Wikipedia-projecten. Alhoewel wij dit wel graag hadden gewild, kunnen we op dit punt aan de hand van de beschikbare statistieken geen eventuele toename van hergebruik over een bepaalde tijd presenteren.

Onze verzameling van meetgegevens is tot stand gekomen op basis van een shortlist van 34 erfgoedinstellingen van Nederland. Voor deze instellingen geldt dat ze alle content beschikbaar stellen op Wikimedia Commons en dat een deel van deze content ook daadwerkelijk wordt hergebruikt op Wikipedia. Onze shortlist werd verkregen door een reeds eerder uitgevoerd onderzoek van OpenGLAM (zie hiertoe het vorige hoofdstuk) onder Nederlandse erfgoedinstellingen. Een aantal van de instellingen die in dat onderzoek voorkomen, hebben ook een dataset op Wikimedia Commons aangeboden. Alleen de Nederlandse erfgoedinstellingen die door BaGLAMa 2 worden vermeld, en ook door deze tool onderzocht kunnen worden, staan op onze shortlist¹⁶. Hier geldt dus een belangrijke restrictie van ons onderzoek: met het totaal van 34 instellingen kan niet aangenomen worden dat dit ook daadwerkelijke alle instellingen zijn die content op Wikimedia Commons aanbieden. Hierover kunnen wij door middel van dit onderzoek geen uitspraken doen. Een voorbeeld hierbij is het Stadsarchief Amsterdam. Deze biedt een OCD-set¹⁷ op Wikimedia aan, maar deze wordt niet door de tool BaGLAMa2 weergegeven en kan dus niet onderzocht worden. Waarom dit precies is, is ons niet duidelijk geworden¹⁸. Via GLAMorous zijn er dan wel uitspraken te doen over het totale aantal documenten op Wikimedia versus hergebruik, echter, de raadpleegstatistieken ontbreken dan¹⁹.

4.5 Meetresultaten voor de Nederlandse aanbieders op Wikimedia

Aan de hand van beide tools hebben we voor iedere instelling de statistieken verzameld. De resultaten van de metingen hebben wij in in een Excel-bestand verwerkt, zie hiertoe bijlage IV. Hierin hebben we de OCD-datasets per instelling bij elkaar gebracht.

Aangezien ons onderzoek zich alleen op de instellingscategorieën bibliotheken, musea en archieven richt, hebben wij in het databestand alle relevante instellingen geselecteerd. Dit maakt dat er een aantal OCD-datasets afvallen. Daarnaast valt ook de “University of Amsterdam Artis Library” af. De content blijkt namelijk verplaatst te zijn naar een andere Wikimedia-categorie, te weten: “Special Collections of the University of Amsterdam”. Het is ons niet duidelijk of alles wat in deze categorie staat van de Artis Bibliotheek is of dat er ook content van andere ‘speciale collecties’ van de Universiteit van Amsterdam hier ondergebracht is. Dit eventueel uifilteren met de tools is onmogelijk, zodat ook deze instelling voor ons onderzoek afvalt.

Voor wat betreft de door het Museum Catharijneconvent aangeboden content geldt dat de metingen pas sinds februari 2015 beschikbaar zijn, zodat dit voor de BaGLAMa 2 tool geen uitkomsten oplevert. Ook voor zeven andere instellingen zijn de metingen pas na december 2014 beschikbaar doordat de instellingen een deel van hun content recenter beschikbaar hebben gesteld. Voor BaGLAMa 2 kunnen deze dus niet meegenomen worden, echter voor GLAMorous wel. Deze instelling blijft dus staan. Hetzelfde geldt voor de instellingen waar wij met GLAMorous geen gegevens over kunnen krijgen.

¹⁶ Uiteindelijk staan bijvoorbeeld wel alle instellingen van het OCD-onderzoek uit 2015 in onze shortlist, maar deze heeft niet als basis gediend.

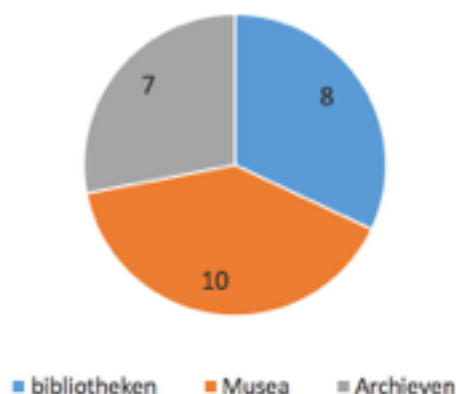
¹⁷ Zie: https://commons.wikimedia.org/wiki/Category:Stadsarchief_Amsterdam

¹⁸ Hiertoe zou inzicht moeten worden verkregen in de exacte (technische) werking van de tool. Dat voert buiten het bestek van ons onderzoek.

¹⁹ In het bestek van dit onderzoek voert het te ver om iedere erfgoedinstelling in Nederland - zowel particuliere als overheidsinstellingen - individueel te gaan benaderen met de vraag of zij een OCD-set op Wikimedia beschikbaar hebben gesteld.

Van de resterende 25 Nederlandse archieven, bibliotheken en musea geeft de onderstaande grafiek weer wat de verhouding daartussen in aantallen is. Op basis van deze grafiek kan worden geconcludeerd dat alle aanbiedende instellingen evenwichtig vertegenwoordigd zijn op Wikimedia.

Verdeling aantal erfgoedinstellingen met hergebruikte open content op Wikipedia



Afbeelding 3. Verhouding archieven, bibliotheken en musea.

Voor wat betreft raadplegingscijfers kan geen nader onderscheid worden gemaakt in type object (afbeelding, video en geluid) aangezien de instellingen die meerdere soorten content aanbieden deze veelal niet hebben uitgesplitst en per type content gevoegd in verschillende categorieën op Wikimedia Commons. Een voorbeeld hierbij is het Rijksmuseum te Amsterdam. De enige uitzondering hier is het Nederlands Instituut voor Beeld en Geluid. Deze instelling heeft binnen één hoofdcategorie²⁰ op Wikimedia Commons vijf OCD-sets gevoegd²¹. Het zijn deze vijf categorieën waar wij met BaGLAMa 2 meer gegevens over hebben kunnen verkrijgen betreffende de raadpleging. De gegevens leveren slechts een vermoeden op voor de rest van de instellingen - hieraan kan geen zinnige conclusie worden verbonden. Overigens blijkt bij een nadere beschouwing dat enkele OCD-sets toch weer ‘vervuild’ zijn geraakt met verschillende soorten content, zoals “Media from Open Beelden” waarin zowel videobestanden als afbeeldingen staan.

4.5.1. Bevindingen op basis van het GLAMorous-onderzoek

Door middel van de tool GLAMorous hebben wij statistische gegevens kunnen analyseren die betrekking hebben op het totale aanbod en het hergebruik van objecten per categorie erfgoedinstelling. Dit levert ons de onderstaande tabel op.

Kolom A	B	C	D	E	F
Categorie instelling	Aanbod: aantal objecten	Hergebruik: aantal Wiki-pagina's met object	Hergebruik: aantal unieke objecten op Wiki-pagina's	% hergebruik van unieke aangeboden objecten (kolom D/B)	Verhouding unieke objecten en aantal Wiki-pagina's met object (kolom C/D)
Archieven	509.804	71.185	26.443	5,2%	2,7
Bibliotheken	7.171	1.864	907	12,6%	2,1
Musea	73.415	43.257	15.803	21,5%	2,7

Tabel 2. Categorieën erfgoedinstellingen: aanbod en hergebruik van (unieke) objecten.

²⁰ Zie: https://commons.wikimedia.org/wiki/Category:Netherlands_Institute_for_Sound_and_Vision

²¹ Zie voor deze OCD-sets tabel 7, instelling no. 3.

In deze tabel is per instellingscategorie het aantal beschikbaar gestelde objecten op Wikimedia Commons – dit is het “aanbod” – afgezet tegen het hergebruik van deze objecten op Wikipedia-pagina’s (kolom B resp. C). De procentuele verhouding tussen het aanbod en hergebruik van objecten is een belangrijke maat voor impact.

Aangezien van ieder object er meerdere kopieën op een of meerdere Wikipedia-pagina’s kunnen voorkomen, is het van belang de unieke objecten uit het totaal van hergebruikte objecten (in kolom C) te filteren. Met behulp van de GLAMorous tool kon dit aantal vastgesteld worden (kolom D) en de verhouding tussen het aanbod en hergebruik worden berekend (kolom E). Aanvullend is bezien of er ook uitspraken kunnen worden gedaan over het aantal kopieën van unieke objecten op een of meerdere Wikipedia-pagina’s. De beoogde gedachte hierbij is dat de impact van een (uniek) object verder toeneemt, naarmate van ieder object meerdere kopieën op Wikipedia circuleren, zoals op de Nederlandse, Duitse of Engelse Wikipedia. De verhouding tussen het totale aantal hergebruikte objecten en het aantal unieke objecten in dat kader is in kolom F in een getal uitgedrukt (kolom C gedeeld door D).

Aan de hand van de door ons gemeten aantallen zijn de percentages in kolom E en de verhouding in kolom F berekend. Vastgesteld kan worden dat bibliotheken voor wat betreft het totale aanbod op Wikimedia Commons duidelijk ondervertegenwoordigd zijn. Musea staan op een tweede plaats en archieven bieden verreweg de meeste content aan.

Voor wat betreft de verhouding tussen het aanbod en hergebruik van unieke objecten op Wikipedia-pagina’s (kolom E) kan worden vastgesteld dat musea het hoogst scoren: 21,5% ten opzichte van bibliotheken (12,6%) en archieven (5,2%). Het voorgaande betekent dat de categorie van musea als mogelijke voorloper aangewezen kan worden. Een andere conclusie die op basis van de metingen kan worden getrokken, is dat het aanbod van objecten niet veel lijkt te zeggen over het uiteindelijke hergebruik van die objecten; archieven bieden namelijk verreweg de meeste objecten aan maar het hergebruik van deze objecten is zeer laag te noemen in vergelijking met bibliotheken en archieven.

Op basis van de waarden in de laatste kolom (F) kan nog vastgesteld worden dat als we kijken naar het gemiddelde van het aantal kopieën van ieder uniek object (die op een of meerdere Wiki-pagina’s kunnen voorkomen), de instellingscategorieën redelijk dicht in de buurt van elkaar scoren: wanneer een object wordt hergebruikt, komt gemiddeld gesproken een bibliotheekobject 2,1 maal voor op een of meerdere Wikipedia-pagina’s en een museaal of archiefobject beide 2,7 maal.

Op basis van de voorgaande bevindingen moet het volgende worden opgemerkt. Wanneer naar het aanbod van archieven wordt gekeken en we daarbij op het niveau van de instellingen zelf inzoomen, valt op dat één instelling verreweg de meeste content aanbiedt, namelijk de Rijksdienst voor het Cultureel Erfgoed: bijna 484.000 afbeeldingen op een totaal van bijna 510.000^{22 23}. Iets soortgelijks kan worden vastgesteld voor musea. Zo biedt het Rijksmuseum Amsterdam circa 23 procent van het totale aanbod van alle musea aan. Het textielmuseum daarentegen biedt maar 17 afbeeldingen aan hetgeen circa 0,02 procent van het totale aanbod is. Ook het aanbod voor wat betreft de bibliotheken is behoorlijk onevenredig te noemen. Derhalve is het maar zeer de vraag hoe valide de voorgaande uitspraken over het hergebruik op het niveau van instellingscategorie zijn. Op basis van het GLAMorous onderzoek luidt onze conclusie dan ook dat er geen voorlopende instellingscategorie aangewezen kan worden.

4.5.2. Bevindingen op basis van het BaGLAMa 2 onderzoek

De tool BaGLAMa 2 analyseert webstatistieken aangaande de raadplegingen (pageviews) van Wikipedia-pagina’s per maand. Aan de hand van onze metingen is de onderstaande tabel samengesteld. Hierin is het

²² Zie bijlage IV, kolom E13.

²³ RCE beheert overigens naast vele archiefstukken ook een bibliotheek waarin bronnen over (bouw-) kunst, bouwhistorie, heemkunde, monumentenzorg, landschap, archeologie, geologie, conservering en restauratietechniek te vinden zijn. Veel van de door RCE beschikbaar gestelde open content is o.i. aan te merken als “archiefstuk”: vandaar dat de OCD-set van het RCE als archief is aangemerkt.

totale aantal raadplegingen per categorie erfgoedinstelling weergegeven. De raadplegingen over het jaar 2014 en 2015 worden naast elkaar weergegeven waarna de procentuele toe- of afname is berekend. De raadplegingen van de objecten op de Nederlandse Wikipediapagina's worden in deze tabel weergegeven.

Categorie instelling	Aantal raadplegingen over 12-2014	Aantal raadplegingen over 12-2015	Toe- of afname in procenten
Archief	35.849.858	65.496.425	183%
Bibliotheek	1.648.842	6.253.969	379%
Museum	21.923.777	31.004.376	141%
Totaal	59.422.477	102.754.770	

Tabel 3. Vergelijking van raadpleegcijfers voor bibliotheken, musea en archieven.

Op basis van deze tabel kan geconcludeerd worden dat er voor alle instellingen een grote toename aan raadplegingen is in één jaar tijd. De toenames in raadpleging voor archieven, bibliotheken en musea zijn 183 %, resp. 379% en 141%. Bibliotheken kennen dus de sterkste toename. Daarnaast zijn bibliotheken sterk ondervertegenwoordigd als het gaat om aantal raadplegingen. Dit zal verband houden met het ook sterk ondervertegenwoordigde totale aanbod van objecten op Wikimedia ten opzichte van musea en archieven. Of er echter een causaal verband vast te stellen is tussen de grote toename in raadplegingen in één jaar tijd en het veel kleinere aanbod van deze instellingen ten opzichte van andere instellingen, valt sterk te betwijfelen. Het verschil tussen musea en archieven qua raadpleging versus totale aanbod laat namelijk precies het omgekeerde zien: dat archieven niet alleen meer content aanbieden maar dat de pagina's waarop deze content terecht is gekomen, ook vaker worden geraadpleegd. Een mogelijke verklaring kan zijn dat de content van met name de bibliotheken encyclopediewaardiger is dan die van musea en archieven.

We zijn van mening dat dit onderzoek te weinig houvast biedt om uitspraken te doen over het bereik van open content van erfgoedinstellingen en kunnen derhalve geen voorloper aanwijzen. Een van de belangrijkste redenen hiervoor is dat, zoals ook het GLAMorous onderzoek al opleverde, de verschillen in aanbod op het niveau van de instelling zeer groot zijn. Door deze onevenredigheid in aanbod, is het zeer lastig uitspraken te doen over gemiddeldes op een hoger aggregatieniveau.

De grootste handicap aan de tool BaGLAMa 2 doet zich voor op het punt dat alleen de raadpleging wordt gemeten van Wikipedia-pagina's waarop content van Nederlandse erfgoedinstellingen staat. Er kunnen geen statistieken verkregen worden die raadpleging van de content zelf aangaat. Dit maakt dat bijvoorbeeld niet vastgesteld kan worden of een gebruiker (rechtstreeks) op zoek was naar een afbeelding of dat hij of zij deze bij toeval op een Wikipedia-pagina tegenkwam. Hierdoor zegt de raadpleging van wikipedia-pagina's feitelijk niet veel over het bereik en daarmee de impact van open cultuur data op Wikipedia.

Om het bereik daadwerkelijk goed te kunnen meten is het onzes inziens van belang dat de beschikbare tools zodanig doorontwikkeld worden dat op basis van de statistieken ten minste kan worden vastgesteld:

- Wat daadwerkelijk als unieke “hit” kan worden aangemerkt voor iedere raadpleging van een pagina (hierover vinden in kringen van ICT-ers voortdurend verwoede discussies plaats);
- Wat de pageviews zijn per individueel object (op Wikimedia Commons), d.w.z.: hoe vaak de objecten daadwerkelijk door Wikipedia-gebruikers bekeken worden;
- Wat de paden zijn die de gebruiker stapsgewijs bewandelt om van een willekeurige Wikipedia-pagina bij de content van een erfgoedinstelling te komen (dus van Wikipedia naar Wikimedia Commons). Het idee hierbij is dat duidelijk gemaakt kan worden welke gebruikers daadwerkelijk content raadplegen

en eventueel ook downloaden. Het laatste geeft ons meer zicht op het verdere hergebruik van open cultuur data en meer inzicht in het informatiezoekgedrag van gebruikers.

4.6. Conclusie

Aan het begin van ons onderzoek hebben wij vast kunnen stellen dat er tools (ofwel: meetinstrumenten) zijn op basis waarvan webstatistieken van Wikipedia geanalyseerd kunnen worden. Specifiek voor de erfgoedinstellingen die op Wikimedia Commons open content en open data voor (her)gebruik aanbieden zijn dat BaGLAMa 2 en GLAMorous. Een eerder onderzoek uit 2015 van Open Cultuur Data in dit kader, waarbij minder instellingen onder de loep werden genomen, leidde tot enkele algemene bevindingen. Met dit onderzoek hebben wij kunnen vaststellen dat analyse van webstatistieken van Wikipedia een behoorlijk braakliggend terrein is. De beschikbare tools BaGLAMa 2 en GLAMorous kunnen behoorlijk de schijn wekken dat de impact van open content en open data rechtstreeks “gemeten” kan worden aan de hand de beschikbare statistieken. Door – wat wij zouden willen noemen – de beperkte mogelijkheden die de tools ons bieden, zijn wij van mening dat er maar tot op zekere hoogte uitspraken kunnen worden gedaan over de impact van Nederlandse open cultuur data op Wikipedia.

Desondanks heeft ons onderzoek wel een aantal bevindingen opgeleverd die iets zeggen over de impact van open cultuur data. Er zijn twee factoren die op impact betrekking hebben én waarover statistische gegevens te verkrijgen zijn, namelijk: hergebruik en bereik. Hergebruik is de mate waarin de door erfgoedinstellingen beschikbaar gestelde content op Wikimedia Commons wordt hergebruikt op Wikipedia-pagina's (zoals die van de Nederlandse en Engelstalige Wikipedia). Bereik is de mate waarin de content geraadpleegd wordt door gebruikers op Wikipedia-pagina's.

Voor ons onderzoek hebben wij vast kunnen stellen dat er ten minste 30 instellingen zijn die een dataset op Wikimedia Commons hebben aangeboden. Daarvan zijn er 25 instellingen aan te wijzen als erfgoedinstelling in de zin van bibliotheken, musea en archieven. Aangezien het niet mogelijk was om ook andere OCD-sets van erfgoedinstellingen op Wikimedia Commons te analyseren dan die op website van de tool BaGLAMa 2 worden vermeld, hebben wij de eerste subvraag van deze paragraaf niet volledig kunnen beantwoorden. Deze vraag luidde immers welke Nederlandse erfgoedinstellingen open cultuur data aanbieden voor hergebruik op Wikimedia Commons. Zelf zijn wij van mening dat het totale aanbod van alle instellingen die een OCD-dataset aanbieden, niet beduidend groter moet worden geacht. De enige juiste manier om een totaalijst samen te stellen is door rechtstreeks in Wikimedia Commons te zoeken naar andere datasets. Hierbij moet opgemerkt worden dat je zelfs ook dan niet over alle content beschikt die gerelateerd kan worden aan de erfgoedinstellingen. Het verrichten van een dergelijke speurtocht zou in het bestek van dit onderzoek en de beschikbare tijd een te grote opgave zijn.

Op basis van de analyses middels de tool GLAMorous hebben wij kunnen vaststellen dat bibliotheken ondervertegenwoordigd zijn als het gaat om het totale aanbod aan objecten. Als naar het hergebruik van deze objecten op Wikipedia wordt gekeken, dan valt op dat musea het hoogst scoren. Daarmee zou deze categorie als voorloper aangewezen kunnen worden. Als echter ingezoomd wordt op de verschillende instellingen, dan valt op dat het aanbod per instelling zeer onevenredig is, en wel zodanig, dat op basis van ons onderzoek geen “voorloper” aangewezen kan worden.

Op basis van de analyses middels de tool BaGLAMa 2 hebben wij slechts zijdelings uitspraken kunnen doen over de raadpleging en dus het bereik van open cultuur data op Wikipedia. Dit onderzoek laat zien dat er over een gemeten periode van december 2014 tot december 2015 er voor iedere instellingscategorie een grote tot zeer grote toename aan raadplegingen van Wikipedia-pagina's (waarop content van Nederlandse erfgoedinstellingen staat) te bespeuren is. Vastgesteld kan worden dat bibliotheken het hoogst scoren. Aangezien ook voor dit onderzoeksgedeelte de grote onevenredigheid in aanbod per instelling in aanmerking moet worden genomen, kan er geen voorloper aangewezen worden. Een

complicerende factor is dat we geen zicht konden krijgen op de raadpleging van de content zelf en het eventueel downloaden ervan, hetgeen een belangrijke gebruikstoepassing kan zijn. De tool is hier – technisch gezien – te beperkt voor.

Ondanks dat ons onderzoek gelimiteerd was tot de OCD-sets van erfgoedinstellingen die door BaGLAMa 2 konden worden geanalyseerd, hebben wij van het bereik en het hergebruik van open content van relatief veel erfgoedinstellingen een redelijk goed beeld kunnen krijgen. Onze eindconclusie luidt dat er geen voorlopers aan te wijzen zijn. Daarmee is de deelvraag van deze paragraaf beantwoord.

5. Gebruik & Gebruikers

5.1 Inleiding

Steeds meer culturele instellingen maken representaties van collectiestukken en informatie over hun collecties, activiteiten en organisatie digitaal beschikbaar. Dit biedt mogelijkheden voor het realiseren van allerlei toepassingen om op nieuwe manieren te kunnen participeren in kunst en cultuur. De aangeboden data is echter vaak slecht toegankelijk. Eindgebruikers kunnen of mogen de informatie maar beperkt (her)gebruiken en de data is technisch slecht benaderbaar. Open Cultuur Data, een gezamenlijk initiatief van Kennisland, Open State Foundation en het Nederlands Instituut voor Beeld en Geluid, stimuleert het openstellen van data uit de culturele sector en het ontwikkelen van nieuwe waardevolle toepassingen, zowel door instellingen zelf als door derden. De data is beschikbaar onder een open licentie of bevindt zich in het Publiek Domein waardoor het iedereen is toegestaan om de data te raadplegen, verspreiden en hergebruiken. De structuur en mogelijke toepassingen van de open cultuurdata worden door de aanbieders gedocumenteerd in de vorm van een datablog op de website (Open Cultuur Data, n.d.).

Open Cultuur Data begon in september 2011 als een netwerk van culturele professionals, ontwikkelaars, designers, specialisten op het gebied van auteursrecht en open data experts met als doel het beschikbaar maken van culturele datasets en het stimuleren van het maken van nuttige en innovatieve toepassingen met deze datasets. Binnen drie maanden na oprichting bevatte Open Cultuur Data acht datasets van zes verschillende aanbieders. Een *hackathon* georganiseerd door Hack de Overheid in november 2011 resulteerde in dertien apps waarvan er acht werden ingezonden naar de ontwikkelwedstrijd Apps voor Nederland en drie apps prijzen in de wacht sleepten. Hoewel het begon als een ad hoc experiment was het duidelijk dat er behoefte was aan openlijk herbruikbare cultuurdata en een sterk netwerk om cultuur te promoten in de (inter)nationale open data beweging. Het Open Cultuur Data-initiatief werd daarom voortgezet en uitgebreid. In 2012 werden er twee activiteiten georganiseerd om data-eigenaren, app-ontwikkelaars en beleidsmakers te verbinden en te inspireren. In het voorjaar vond er een masterclass van Creative Commons Nederland plaats, waar vertegenwoordigers van verschillende GLAM-instellingen lezingen gaven over het proces van het open toegankelijk maken van culturele datasets. Zeventien nieuwe aanbieders van cultuurdata namen deel aan deze masterclass waarin er onder andere werd gesproken over risico's en voordelen van open data. Vervolgens werd er een app-competitie opgericht met als doel het stimuleren van ontwikkelaars om de open cultuur datasets op een creatieve manier te hergebruiken. Er werden voor deze competitie 27 nieuwe apps ontwikkeld, die echter vaak weer verdwenen omdat ze niet op de lange termijn konden worden ondersteund en niet verder kwamen dan de prototype-fase. Investeren in meer duurzame toepassingen is een belangrijke les voor de toekomst die voortkwam uit de app-competitie (Baltussen et al.).

Op dit moment omvat Open Cultuur Data meer dan 60 datasets van 37 verschillende culturele instellingen (zie bijlage III) . Het is moeilijk vast te stellen wie de daadwerkelijke gebruikers van de data precies zijn. Het is wel mogelijk om aan de instellingen te vragen of zij weten wie hun gebruikers zijn of wat het beoogde gebruiksdoel van de datasets is, om zo toch inzicht te krijgen in mogelijkheden voor gebruikers. Om dat te bereiken is een e-mailactie met een aantal concrete vragen over open data gebruik, gebruikers en voorbeelden van toepassingen verzonden naar 36 Open Cultuur Data-instellingen. Arts Holland is niet benaderd omdat dit een afgerond project betreft. Deze e-mailactie heeft 17 reacties opgeleverd (zie bijlage x voor de email tekst en ruwe data).

5.2 Bevindingen

Deelvraag 3:

Wie zijn de gebruikers van open data, wat zijn de gebruikstoepassingen en hoe wordt gebruik van open data gemeten? Hieronder wordt per vraag uiteengezet wat de bevindingen zijn.

De antwoorden op vraag 1 t/m 4 zijn in tabellen gezet. Bij vraag 5 zijn de antwoorden dusdanig kwalitatief dat het niet mogelijk is de gegevens in een tabel te verwerken.

1. Door wie of door welke instellingen wordt uw Open Data volgens u gebruikt?

(bv. particulieren / onderwijs / overheid / culturele instellingen / ondernemingen / onderzoekers / media / onbekend / andere, namelijk ..)

Vraag 1	particulieren	onderwijs	overheid	culturele instellingen	ondernemingen	onderzoekers	media	onbekend	anders
Amsterdam Museum	x	x		x	x	x		x	
Archief Vlissingen	x	x		x		x			
Beeld en Geluid	x	x	x	x	x	x	x		
Catharijneconvent	x	x				x	x		
Comité 4 en 5 mei	x	x	x	x		x	x		
Gemeente Rotterdam	x								
Open archieven						x			
Regionaal archief Alkmaar	x								
Regionaal archief Nijmegen	x		x			x	x	x	
Regionaal archief Tilburg	x				x	x			
Rijksdienst voor het Cultureel Erfgoed				x	x				Catalogi
Rijksmuseum van Oudheden								x	
Rijksmuseum	x	x			x	x			Portal sites
Textielmuseum		x		x					OCD
Tropenmuseum	x			x		x			source communities
Universiteitsbibliotheek Utrecht								x	
NIOD Instituut voor Oorlogs- Holocaust en Genocidestudies	x	x		x		x	x		
Totaal	12	8	3	8	5	11	5	4	schrijvers

Wat opvalt is dat 12 van de 17 instellingen aangeven dat vooral particulieren en onderzoekers gebruikmaken van hun open data. De datasets die deze instellingen aanbieden omvatten met name beeldmateriaal en genealogische data (zie bijlage x overzicht OCD datasets). De portal Open Archieven vertekent het beeld enigszins. Zij geeft aan de volgende typen gebruikers te hebben:

- technisch georiënteerde (stamboom)onderzoekers;
- internationale genealogische bedrijven (als Ancestry.com en MyHeritage.com);
- (informatie/geschiedenis)studenten.

Geen particuliere gebruikers dus, maar via de bij het tweede punt genoemde websites is de data gewoon beschikbaar en kan dus ook door particulieren geraadpleegd worden. Media, ondernemingen en de overheid blijven over de gehele linie achter als het gaat om gebruik. Iets wat daartegenover helemaal niet verbaast is dat met name musea het onderwijs als gebruikers zien.

Het is verder nodig om een nuancering aan te brengen met betrekking tot de doelgroep genealogen die veelal onder onderzoekers wordt geschaard. Het is namelijk niet altijd duidelijk of het hier om hobbyisten of wetenschappers gaat. Daarom wordt deze groep soms in de categorie particulieren en soms in de categorie onderzoekers ondergebracht. Dit hangt af van de informatie die de instelling over deze gebruikersgroep geeft. Het Tropenmuseum geeft bijvoorbeeld als antwoord: “Onderzoekers, waaronder professionals uit de

museum- en academische wereld, liefhebbers, genealogen, museumbezoekers, partnermusea, source communities”. Zowel onderzoekers als particulieren (liefhebbers) worden hier al apart genoemd, genealogen kunnen onder beide noemers vallen.

Het Comité 4 en 5 mei geeft een uitgebreide weergave van de gebruikers van diens open data: “Culturele instellingen (zoals regionale of lokale erfgoed partijen en organisaties die een herdenking of viering organiseren), overheid (ministerie van VWS, provincies, gemeenten, RCE), onderwijs (basisscholen in het kader van ons project “Adopteer een monument”), onderzoekers, media, algemeen geïnteresseerden en onze data wordt veel gebruikt in veelal regionale en lokale publicaties van (amateur)historici”.

2a en b. Welke toepassingen denkt u dat er met uw Open Data mogelijk zijn?

Vraag 2a	Apps voor mobiele telefoon of tablet	Apps voor mobiele telefoon of tablet ontwikkeld op basis van geografische data en/of beeldmateriaal	Apps voor mobiele telefoon of tablet waarmee een catalogus kan worden geraadpleegd	Apps voor mobiele telefoon of tablet in de vorm van games	Apps voor mobiele telefoon of tablet gebruikmakend van audiovisueel materiaal (remixen van materiaal, samenstellen van een eigen collectie)	overig
Amsterdam Museum		x	x	x		
Archief Vlissingen						
Beeld en Geluid	x	x	x	x	x	
Catharijneconvent	x	x		x	x	
Comité 4 en 5 mei	x	x		x		
Gemeente Rotterdam		x				
Open archieven						
Regionaal archief Alkmaar	x	x	x	x	x	
Regionaal archief Nijmegen	x	x	x	x	x	
Regionaal archief Tilburg						
Rijksdienst voor het Cultureel Erfgoed			x	x	x	
Rijksmuseum van Oudheden						
Rijksmuseum	x	x		x	x	screensavers; commerciële apps
Textielmuseum	x	x	x			
Tropenmuseum						
Universiteitsbibliotheek Utrecht						
NIOD Instituut voor Oorlogs-Holocaust en Genocidestudies		x		x	x	
Totaal	7	10	6	9	7	

Vraag 2b	Open Data elders op het web buiten uw instelling opvraagbaar en herbruikbaar	Gebruik in wetenschappelijk onderzoek, educatieve toepassingen, lesmateriaal	Open Data delen via Social Media	Gebruik in reclame en/of media	Overige toepassingen
Amsterdam Museum	x	x	x	x	
Archief Vlissingen		x			
Beeld en Geluid	x	x	x	x	
Catharijneconvent		x	x		Wikicomments
Comité 4 en 5 mei	x	x	x		
Gemeente Rotterdam	x	x			
Open archieven		x			
Regionaal archief Alkmaar	x	x	x	x	
Regionaal archief Nijmegen	x	x	x	x	
Regionaal archief Tilburg	x	x			
Rijksdienst voor het Cultureel Erfgoed			x	x	
Rijksmuseum van Oudheden					
Rijksmuseum	x	x	x	x	
Textielmuseum	x	x	x		Tijdlijn/kaarten
Tropenmuseum					
Universiteitsbibliotheek Utrecht					
NIOD Instituut voor Oorlogs-Holocaust en Genocidestudies	x		x	x	
Totaal	10	12	10	7	

De meeste instellingen hebben de vraag geïnterpreteerd als “Welke toepassingen maken gebruik van uw data?”. Dit heeft invloed op de eenduidigheid en daarmee de validiteit van de antwoorden op deze vraag. Vraag 2 is in een a en b deel uitgesplitst om de data makkelijker te kunnen verwerken. Deel a gaat over specifieke app toepassingen. Deel b gaat over andere toepassingen die ook als optie waren gedefinieerd in de vragenlijst.

Het valt op dat bij vraag 2a 6 van de 17 instellingen niets hebben ingevuld. Bij 2b zijn er 3 van de eerder genoemde 6 die weer niets invullen. De catalogus-apps worden met name bij archieven genoemd en minder bij Musea wat niet in de lijn der verwachting ligt. Musea hebben immers meer beeldmateriaal beschikbaar dat zich goed leent voor exploratie in een webomgeving. 12 van de 14 instellingen die iets ingevuld hebben bij 2b geven aan dat hun data gebruikt wordt voor wetenschappelijk onderzoek, educatieve toepassingen en lesmateriaal.

Archief Vlissingen geeft als mogelijke toepassingen aan: “Gebruik voor historici en dan niet alleen maritiem georiënteerd, maar bijvoorbeeld ook landbouw, educatieve toepassingen op school voor bijvoorbeeld de industriële revolutie maar ook in het technisch onderwijs. Modelbouwers”. De gemeente Rotterdam laat weten: “Op dit moment bestaan onze open datasets voornamelijk uit sets met genealogische data. En we hebben open dataset met historische straatnamen. De genealogische data die we beschikbaar hebben gesteld zijn goed bruikbaar voor wetenschappelijk onderzoek en om deze data elders op het web buiten onze eigen website opvraagbaar en herbruikbaar te maken. De straatnamen-dataset kan gebruikt worden voor apps op basis van geografische data”.

3. Welke concrete voorbeelden kunt u geven van gebruikers van uw Open Data en/of gebruik van uw Open Data in toepassingen?

Vraag 3	Gebruik	Gebruikers
Amsterdam Museum		Onderzoekers bij de UvA en de Vu
Archief Vlissingen		
Beeld en Geluid	Wikipedia	
Catharijneconvent	Wikipedia	Facebook
Comité 4 en 5 mei		
Gemeente Rotterdam	Openarch.nl, erfgeo.nl	
Open archieven	Ancestry.com, myheritage.com	Informatica student gebruikt het voor zijn bachelorscriptie
Regionaal archief Alkmaar	Genealogische data via Open Archieven. EAD's via Archives Portal Europe	Blogster via Flickr
Regionaal archief Nijmegen	Beeld en Geluid, Volkskrant, Wikipedia, Open Archieven, Europeana, Collectie Gelderland	Ontwerpers
Regionaal archief Tilburg	Openarchieven.nl, myheritage.nl	
Rijksdienst voor het Cultureel Erfgoed	Europeana, wikimedia, Virtuele Fachbibliotheek	
Rijksmuseum van Oudheden		
Rijksmuseum		
Textielmuseum	Portal opencultuurdata.nl	Onderwijsinstellingen
Tropenmuseum	Wikipedia	
Universiteitsbibliotheek Utrecht	webkaart van provincie Utrecht	
NIOD Instituut voor Oorlogs- Holocaust en Genocidestudies	Oorlogsbronnen.nl, Europeana	

Bij deze vraag gaat het vooral om aanvullende informatie op vraag 1 en 2. Hieronder wordt een aantal interessante voorbeelden uit de gegeven antwoorden gelicht. Open Archieven geeft aan: “Ancestry.com en MyHeritage.com hebben de data geharvest en bieden zogenaamde Record Matches aan, zij wijzen hun gebruikers dus ook op bronnen gelinkt aan personen in hun stamboom”. Hierbij wordt bevestigd wat bij vraag 1 al aan de orde is gekomen. Zeer divers zijn bijvoorbeeld de toepassingen die Regionaal Archief Nijmegen aangeeft:

- Verwerken posters poppodium Doornroosje in tentoonstelling in en door Beeld en Geluid in Hilversum;

- Digitale remixes van posters van Doornroosje door ontwerpers;
- Publicatie in de Volkskrant van selectie foto's bevrijders voor de lens;
- Plaatsing op Wikimedia/Wikipedia van Vierdaagse foto's door Wikipedianen;
- Gebruik woningkaarten voor crowdsourcing project op Open Archieven;
- Metadata beeldcollectie (en beelden waar het auteursrechtelijk kan) beschikbaar via portals als Europeana en CollectieGelderland.

Het Rijksmuseum is één van de vijf instellingen die aangeeft dat hun data redelijk veel gebruikt wordt in reclame en/of media. Zij geven de volgende voorbeelden: “zowel met als zonder onze samenwerking. Soms heel leuk: Victor en Rolf collectie, La Dress, Eigen Huis en Tuin, Melk van de Albert Heijn enz.”

Tropenmuseum: “Tropenmuseum (of correcter: collectie van het Nationaal Museum van Wereldculturen, maar tot nu toe is het allemaal Tropenmuseum collectie) heeft bijna 50.000 beelden geupload naar Wikimedia commons. Die beelden worden ter illustratie van lemma's in Wikipedia gebruikt, en dat resulteert in 10-30 miljoen pageviews (van de lemma's dus) wereldwijd, per maand. Of ze daarna via Wikimedia de weg naar onze website vinden is een tweede, maar dat is niet het primaire doel. Feit is: content wordt gebruikt door de Wiki-community én gezien door de gebruikers van Wikipedia.

De informatie die op onze online collectie te zien is mag vrij gebruikt worden. Beelden zijn in lage resolutie te downloaden, dat houden we niet actief tegen. Een disclaimer hebben we op het moment niet, hoewel er beeldmateriaal te downloaden is waarvan wij de rechten niet hebben. Dat is dan echter wel vermeld.”

4. Houdt u bij wat er van uw Open Data wordt gebruikt, hoe vaak dat is en waar? Zo ja, waarmee kunt u dat meten?

Vraag 4	Houdt u gegevens bij?					Waarmee?	In ontwikkeling?
	Houdt u wel gegevens bij (geaggregeerd)	Actief bijhouden (handmatig)	Passief bijhouden (beslaat het gebruik dat direct aan de instellingen wordt gerapporteerd door derden)	Webmetrics (Google Analytics, Wikipedia tools)	Houdt geen gegevens bij		
Amsterdam Museum	WAAR	x				Handmatig via Google	
Archief Vlissingen	WAAR			x		Wikipedia en Facebook	
Beeld en Geluid	WAAR		x	x		Wikipedia	
Catharijneconvent	WAAR			x		Wikipedia en Facebook en eigen website	
Comité 4 en 5 mei	ONWAAR				x		x
Gemeente Rotterdam	WAAR		x			Via Open Archieven	
Open archieven	WAAR			x		Google analytics en via de eigen website	
Regionaal archief Alkmaar	WAAR	x	x	x		Flickr (views niet downloads), wikipedia, Google analytics, via Open Archieven	
Regionaal archief Nijmegen	WAAR	x					
Regionaal archief Tilburg	WAAR		x		x		
Rijksdienst voor het Cultureel Erfgoed	WAAR		x		x		
Rijksmuseum van Oudheden	ONWAAR				x		
Rijksmuseum	WAAR	x	x	x		Wikipedia, europeana	
Textielmuseum	WAAR		x			Niet gespecificeerd	
Tropenmuseum	WAAR			x		Google analytics indirect via wikipedia en europeana wordt niet bijgehouden	
Universiteitsbibliotheek Utrecht	ONWAAR				x		
NIOD Instituut voor Oorlogs-Holocaust en Genocidestudies	ONWAAR				x		
TOTAAL		0	4	7	7	6	1

Veel instellingen gebruiken de webmetrics van open data voor het relativeren van hun publieke functie. Onder het motto “meten is weten” is deze vraag voorgelegd aan de instellingen om inzicht te verkrijgen in het gebruik van deze statistieken door de instellingen. Met het oog op de respons van de instellingen zijn er een aantal kanttekeningen te plaatsen bij de interpretatie van deze vraag. De vraag die de instellingen lijken te hebben beantwoord is als volgt: Meet u op één of andere manier het gebruik van de open data en zo ja, waarmee? Het bespreken van de andere twee onderdelen van de vraag: hoe vaak en wat wordt er van uw Open Data gebruikt, wordt hierbij achterwege gelaten.

Een belangrijk onderdeel van deze vraag dat enige uitleg behoeft is het gebruik van open data. Wat is precies gebruik, en wellicht relevanter: “Wat verstaan de bevraagde instellingen onder gebruik?” Open Archieven zegt hier het volgende over: “De data wordt geboden via een API (JSON) en via een OAI-PMH dataprovider. Elk request wordt als een event gelogd via Google Analytics. Maar dit zegt natuurlijk niet zo veel over het gebruik van de data. Elke keer als er een Record Match door MyHeritage wordt gedaan zou je bijv. moeten tellen als één "gebruik". Maar dat alle data door de Informatica student worden verwerkt in inzichtvolle (sic) grafieken, hoeveel "gebruik" is dat?” Andere instellingen benoemen dit niet, waardoor de respons in veel gevallen multi-interpretabel is. Bijvoorbeeld het Amsterdam Museum geeft aan dit soort informatie enkel passief bij te houden, er wordt met een zoekopdracht via Google bekeken op wat voor manieren de open data van het museum wordt gebruikt. Hierbij is aangenomen dat de respondent verwijst naar daadwerkelijk (her)gebruik en niet naar gebruiksstatistieken.

Bij het verwerken van de respons is ook een onderscheid gemaakt tussen het actief en passief bijhouden van dit soort informatie. Onder actief wordt het handmatig bijhouden van het gebruik verstaan. Passief beslaat de gebruiksstatistieken die direct aan de instellingen worden gerapporteerd of direct opvraagbaar zijn. Dit kan bijvoorbeeld via Google Analytics, Flickr, Facebook of de Wikimedia tools genoemd in hoofdstuk 4, of door de ontwikkelaar die contact opneemt met de instelling.

13 van de 17 instellingen geven aan informatie over gebruik bij te houden. 7 van de 13 instellingen zeggen met name afhankelijk te zijn van informatie van derden (bijv. gebruikers van de data) over gebruik. Dit overlapt deels met de instellingen die een combinatie van middelen inzetten om het gebruik van de open data voor de instellingen inzichtelijk te maken.

7 van de 17 instellingen gebruiken webometrics om (her)gebruik van de collectie bij te houden. 5 van deze 7 instellingen interpreteren gebruik in de context van de vraag dan ook als zijnde gebruiksstatistieken, deze instellingen aggregeren hiervoor statistieken vanuit meerdere platformen of systemen. Informatie van bijvoorbeeld Google Analytics en Wikimedia Commons moet dan worden samengevoegd om een geheel ‘plaatje’ te schetsen van het (her)gebruik van de opengestelde data. Dit is met name omdat views binnen de verschillende platformen misschien anders worden geteld. Zoals het Regionaal Archief Alkmaar ook al aangeeft:

“We proberen alles bij te houden, maar niet iedereen geeft het door als ze het gebruiken. Flickr laat bijvoorbeeld alleen views van de beelden zien, geen downloads. Zelfde geldt voor Wikimedia. Wel meten we via Google Analytics wie er via verschillende bronnen op onze site komen. Open Archieven geeft ook inzicht in hoe vaak er views zijn van onze genealogische data op hun site. Maandelijks rapportage naar onze directie.”

5. Wat hoopte u te bereiken door Open Data beschikbaar te stellen? Zijn uw verwachtingen uitgekomen?

Verschiedende instellingen geven aan dat het vergroten van de naamsbekendheid en van het bereik van de collectie de belangrijkste doelen waren van het beschikbaar stellen van open data. Deze organisaties zijn over het algemeen tevreden over de behaalde resultaten tot dusver.

Het Regionaal Archief Alkmaar laat weten: “We hopen onze collecties en naam een groter bereik te geven. Dat is wel gelukt. Vanaf het begin waren we niet in de veronderstelling dat er nu massaal gebruikt zou gaan worden gemaakt van onze data. Dat blijkt ook zo te zijn geweest (in 2015)”. Het Catharijneconvent: “Verspreiding van de collectie en vergroten van de naamsbekendheid. Zijn heel tevreden”. De Rijksdienst voor het Cultureel Erfgoed, het Textielmuseum en het NIOD geven vergelijkbare antwoorden.

Instellingen die verwachtingen hadden over concreet gebruik van de data zijn niet allemaal content, zij hadden soms de hoop dat er meer toepassingen zouden worden ontwikkeld. Zo geeft de Universiteitsbibliotheek Utrecht aan: “We hoopten dat slimme ontwikkelaars apps zouden gaan bouwen waarmee de collectie oude kaarten bij de Universiteitsbibliotheek Utrecht in bezit meer zichtbaarheid zou krijgen. Onze verwachtingen zijn tot dusver helaas niet geheel uitgekomen.” De contactpersoon van Archief Vlissingen zegt: “De verwachtingen zijn niet geheel uitgekomen. Het was een uniek project en we hadden een veel groter gebruik verwacht. Misschien dat het in het Nederlands is en niet in het Engels. Via weblogs, LinkedIn, Facebook is er wel geregeld naar verwezen.”

Het Rijksmuseum daarentegen is zeer enthousiast: “Onze visie te delen dat wij open en van iedereen zijn. Het komt voort uit een principiële visie van het museum en heeft onze stoutste verwachtingen overtroffen: onze collectie is echt ‘out there’. Mensen kennen ons en weten dat we hét Rijksmuseum zijn. Bovendien behoren we inmiddels tot de mondiale voorlopers op gebied van digitale openheid. Dat hadden we nooit verwacht of bedacht... En onze collectie leeft en wordt gedeeld en inspireert. En daar deden we het voor”. Ook het Tropenmuseum is tevreden over het gebruik van de data. Het gaat hier om twee grotere musea die waarschijnlijk een groter bereik hebben dan de meeste andere culturele instellingen.

Beeld en Geluid, de Gemeente Rotterdam, het Regionaal Archief Nijmegen, het Regionaal Archief Tilburg en het Amsterdam Museum zijn vooralsnog positief over het open data experiment en benieuwd naar wat de toekomst nog gaat brengen. Het Comité 4 en 5 mei en het Rijksmuseum van Oudheden hadden geen concrete doelen of verwachtingen voor ogen, buiten het voor iedereen beschikbaar stellen van de data.

5.2 Conclusie

In dit hoofdstuk is getracht drie vragen te beantwoorden. Allereerst wie de gebruikers zijn van open data, ten tweede wat de gebruikstoepassingen zijn en tot slot hoe het gebruik van open data wordt gemeten. Met betrekking tot de eerste deelvraag geven zowel archieven als musea vergelijkbare antwoorden. De instellingen zien particulieren en onderzoekers als voornaamste gebruikers van de open data, met culturele instellingen en het onderwijs op een gedeelde derde positie. Dit is ook nog een keer terug te zien in vraag 2b, waar gebruik van de open data bij wetenschappelijk onderzoek en als onderdeel van educatieve toepassingen door vrijwel alle (12) instellingen wordt aangevinkt als mogelijkheid. Overige gebruikersgroepen blijven wat achter, met name de overheid.

Verdergaand met de respons van vraag 2b zien we dat "open data elders buiten de instelling opvraagbaar en herbruikbaar" en "open data delen via social media" als belangrijke mogelijkheden worden herkend. Met het oog op de verwachtingen in vraag 5, waar het vergroten van het bereik van erfgoedmateriaal als erg belangrijk werd/wordt ervaren, is dit dan ook niet verbazingwekkend.

13 van de 17 instellingen zeggen op één of andere manier het gebruik van open data bij te houden. Er is een onderscheid gemaakt tussen drie verschillende manieren waarop dat door instellingen lijkt te gebeuren, achter de methode staat tussen haakjes het aantal instellingen die aangeven deze methode te gebruiken. Het bijhouden van gebruik gebeurt dan wel handmatig (4), waarbij via Google wordt gezocht naar open data van de instelling, een andere manier is passief (7), waarbij de instelling door derden op de hoogte wordt gehouden van gebruik. Webometrics wordt als laatste categorie genoemd (7), hierbij worden automatisch statistieken omtrent gebruik verzameld via bijvoorbeeld Google Analytics of de tools van Wikimedia Commons. Vijf instellingen geven aan in het geheel geen gegevens bij te houden.

Als kanttekening kan aangemerkt worden dat een aantal van de vragen multi-interpretabel zijn. Soms wordt de vraag helemaal niet begrepen door de instelling. Het gevolg kan zijn dat de validiteit van de antwoorden op de andere vragen hierdoor wordt beïnvloed. Bij een aantal van de vragen lijkt met name de woordkeuze voor problemen te zorgen. Het gebruik van open data bijvoorbeeld. Gebruik is in deze context

geen vastliggend begrip, zijn het views, downloads of is het met name hergebruik? Uit de antwoorden is niet altijd op te maken wat de instelling hieronder verstaat.

De instellingen die de vragenlijst hebben ingevuld zijn over het algemeen tevreden over het open data-experiment. Enkele aanbieders hadden meer gebruik van de data verwacht maar andere zijn al tevreden met een groter bereik van de data en een grotere naamsbekendheid en zijn benieuwd naar wat de toekomst nog brengt op het gebied van open data.

6. Conclusie

In het onderzoek is ingegaan op de centrale vraag: ‘Welke aspecten spelen een rol bij het aanbieden van open data van Nederlandse erfgoedinstellingen op dit moment en in de nabije toekomst en wat is de impact van het gebruik van open data?’

Deze vraag is verder in drie deelvragen uitgewerkt en de betreffende hoofdstukken 3, 4, en 5 bevatten elk gedetailleerde specifieke deelconclusies die de moeite van het lezen waard zijn en daarom in deze samenvattende conclusies niet integraal herhaald worden. Een aantal bevindingen uit de deelonderzoeken zijn hieronder echter wel met elkaar in verband gebracht om kanttekeningen te kunnen plaatsen bij het uitgevoerde onderzoek en aanbevelingen te kunnen geven voor toekomstig onderzoek.

De antwoordgegevens uit het OpenGLAM onderzoek van Estermann zijn volgens ons zeker te gebruiken voor een nadere analyse van de adoptie van open data voor erfgoedinstellingen in Nederland, maar geven slechts een beperkt beeld van de achterliggende motivaties van instellingen om aan open data bij te dragen. Ook blijft het onduidelijk wie de gebruikers van open data zijn en hoe ver de impact van open data reikt. Om dat te kunnen onderzoeken zijn aanvullende methoden nuttig. In dit onderzoek hebben de bestudering van twee specifieke tools om Wikidata te meten en een enquête onder erfgoedinstellingen, waarvan bekend is dat zij open data beschikbaar hebben, waardevolle aanvullingen laten zien voor het debat over de rol van erfgoedinstellingen bij het beschikbaar stellen van open data en de impact ervan.

In het OpenGLAM onderzoek zijn weliswaar uitspraken gedaan over verbanden tussen adoptie van open data en karakteristieken van erfgoedinstellingen, maar die gaan vaak voorbij aan de situatie per land en per instellingstype (bijvoorbeeld archieven of musea). Als we de onderzoeksdata voor Nederland opnieuw bekijken en analyseren dan blijkt een aantal uitspraken niet van toepassing te zijn of genuanceerd te moeten worden. Zo is er weliswaar een verband tussen instellingsgrootte en adoptie van open data, maar dat geldt specifiek in Nederland alleen voor archieven (niet voor musea) en ook alleen voor open metadata (niet voor open content). Het verband dat Estermann suggereert tussen adoptie van open content en het feit dat instellingen aan Wikidata bijdragen wordt door ons voor Nederland niet gevonden. Tal van detailvragen kunnen nog m.b.v. de beschikbare onderzoeksdata worden gesteld en statistisch onderzocht, en er is nog veel werk mogelijk op dit gebied. We hebben geprobeerd hiertoe een aanzet en bijdrage te leveren. Overigens zijn alle bevindingen gebaseerd op een steekproef van 146 instellingen, terwijl bijna veertienhonderd instellingen voor het OpenGLAM onderzoek waren aangeschreven. De vraag blijft in hoeverre deze steekproef representatief is. Er is wel een evenwichtige verdeling tussen archieven en musea en in instellingsgrootte geconstateerd, die vertrouwen geeft in de samenstelling van de steekproef.

In detail kijkend naar een aantal instellingen die aan het OpenGLAM onderzoek hebben meegedaan ontstaat het gevoel dat sommige instellingen minder ver zijn met het beschikbaar stellen van open data dan door hen zelf aangegeven wordt. Een blik op hun website roept in deze gevallen het beeld op van een minder professioneel karakter en minder gevorderde technische equipage. Het lijkt erop dat men het tonen van afbeeldingen op de eigen website van erfgoedobjecten al een vorm van open data vindt. De manier waarop in het OpenGLAM onderzoek is uitgelegd wat open data precies is kan door instellingen onvoldoende begrepen worden, en zo kunnen bepaalde vragen over dit onderwerp onbedoeld onzorgvuldig worden ingevuld. Ook om deze reden is de aanvullende enquête die binnen dit onderzoek is uitgevoerd interessant om iets meer te weten te komen over hoe instellingen denken over open data, hun motivatie en hun kennis omtrent gebruikers en gebruik. Uit de variatie in antwoorden op de vraag naar toepassingen van open data en het gebruik blijkt ook daar veel verschil in definitie van open data en de opvattingen over mogelijke toepassingen.

In het OpenGLAM onderzoek wordt niets gevraagd over het gebruik van open data en het meten van impact. Omdat we uit dat onderzoek wel een verband hebben gevonden tussen adoptie van open data en het

feit dat veelal publieke erfgoedinstellingen met publieke gelden gefinancierd worden, mag verwacht worden dat er ook behoefte bestaat aan het meetbaar maken van het bereik en de mate van (her)gebruik van open data omdat aan het publiek kenbaar te maken. Ook uit de door ons uitgevoerde enquête blijkt deze behoefte te bestaan. Het merendeel van de instellingen gebruikt beschikbare statistieken van Google Analytics, Facebook, Flickr en Wikimedia naar hartenlust in combinatie, aangevuld met statistieken door derden. Ook door te googlen naar de eigen open data van de instelling wordt informatie over hergebruik verzameld. Naar beste vermogen wordt zo door de instellingen een eigen beeld van impact gevormd op basis van onvergelykbare metingen. Uit de enquête blijkt ook dat er verschil van interpretatie is als het gaat over gebruik en het meten van open data. Het is niet altijd duidelijk wat een instelling hieronder verstaat.

Voor het onderzoek naar instrumenten (tools) voor impactmeting van open content en data op Wikipedia kan het volgende worden vastgesteld. Impact kan aan de hand van de factoren hergebruik en bereik worden gemeten waarbij hergebruik de mate is waarin de door erfgoedinstellingen beschikbaar gestelde content wordt hergebruikt op Wikipedia en bereik een maat voor raadpleging. Hiervoor dienen inzichtelijke en bruikbare meetinstrumenten ingezet te worden. Specifiek voor de erfgoedsector zijn er slechts twee zogeheten GLAMetrics-tools beschikbaar: BaGLAMa 2 en GLAMorous. Deze tools hebben wij op doelmatigheid onderzocht. De BaGLAMa 2 en GLAMorous tools zijn veelbelovend, maar meten slechts een deel van de beschikbaar gestelde en uiteindelijk hergebruikte open cultuur data. Er kunnen op basis van de geanalyseerde webstatistieken geen uitspraken gedaan worden over de reden van raadpleging en hergebruik. We kunnen hooguit uitspraken doen over hoe vaak Wikipedia-pagina's worden geraadpleegd waarop content van de Nederlandse erfgoedinstellingen staat. De raadpleging zegt niets over hoe vaak die objecten worden geraadpleegd en eventueel ook gedownload, hetgeen een belangrijke gebruikstoepassing zou kunnen zijn. Wel kan er worden vastgesteld welk deel van de aangeboden open content op Wikimedia Commons wordt hergebruikt op Wikipedia-pagina's en om hoeveel unieke objecten het daarbij gaat. Daarmee kunnen archieven, bibliotheken en musea met elkaar worden vergeleken. Op dit punt laat ons onderzoek zien dat de content van musea gemiddeld vaker wordt hergebruikt dan de content van archieven en bibliotheken. Het feit dat bibliotheken ten opzichte van archieven en musea verhoudingsgewijs zeer weinig content aanbieden en er daarnaast ook zeer grote verschillen in aanbod per individuele instelling zijn te bespeuren, maakt dat bibliotheken niet als "voorloper" aan te merken zijn. Voor beide tools geldt dat er geen duidelijke methoden beschikbaar zijn om voorlopers van open content te kunnen karakteriseren. Een belangrijke bevinding is dat de GLAMetrics-tools die hier zijn onderzocht weliswaar iets van inzicht geven in het gebruik en bereik van open content, maar dat de tools hiervoor verder ontwikkeld moeten worden. Onderzoek op meerdere terreinen van open data is nodig om als volwaardige meetinstrumenten voor impactanalyse te kunnen fungeren. Het ontbreekt met name ook aan metingen in de tijd, waardoor trends en prognoses mogelijk worden. Hier is nog veel werk te doen voor computer- en informatiewetenschappers.

Om op een andere manier meer inzicht te krijgen in de gebruikers en het gebruik van open data is in dit onderzoek een aanvullende enquête gedaan met een vijftal vragen onder 36 erfgoedinstellingen die hebben bijgedragen aan het OpenCultuurData initiatief. Deze enquête, per email en met zo doelgericht mogelijke mailadressen, heeft binnen een week een goede respons opgeleverd door zeventien instellingen. De ondervraagde instellingen laten een grote variatie zien in gebruikersgroepen van hun open data, maar het valt op dat vrijwel alle erfgoedinstellingen aangeven dat zij particulieren als belangrijke gebruiker zien. Data die wordt aangeboden door deze instellingen bestaat vooral uit beeldmateriaal (open content) en genealogische data (open metadata). Onderzoekers scoren ook hoog als gebruiker van open data, gevolgd door onderwijs en culturele instellingen. Aan een vergelijking tussen de resultaten uit OpenGLAM, op het punt van de gebruikers van erfgoedinstellingen, met de resultaten uit de eerste enquêtevraag is in dit onderzoek niet toegekomen. Dat zou per instellingstype (archieven en musea) een interessante vergelijking kunnen zijn omdat niet dezelfde instellingen zijn ondervraagd en er ruim een jaar is verstreken tussen beide

onderzoeken. In het OpenGLAM onderzoek is weliswaar gevraagd aan instellingen of zij open data belangrijk vinden en kansrijk, maar de werkelijke redenen blijven onbelicht. Daarom is in de enquête gevraagd naar de motivatie van de instelling om open data beschikbaar te stellen en of de verwachtingen zijn uitgekomen. Dit geeft een idee over de toekomst van open data en of er groei is te verwachten. Belangrijke doelen die worden aangegeven zijn het vergroten van de naamsbekendheid en het vergroten van het bereik van de collectie. De meeste instellingen geven aan daarin geslaagd te zijn. Instellingen zien mogelijkheden om open content buiten de instelling beschikbaar te maken en te delen via sociale media. Dit sluit goed aan bij voorgaande doelstellingen. Over concreet gebruik en hergebruik van open data zijn de verwachtingen minder unaniem, het zijn vooral grote instellingen als het Rijksmuseum en Tropenmuseum die er in zijn geslaagd het gebruik te promoten en meten.

Afrondend stellen we vast dat het niet eenvoudig is om erfgoedinstellingen te karakteriseren die voorop lopen in de adoptie van open data op basis van alleen de gegevens verkregen uit het OpenGLAM onderzoek. Andere onderzoeksmethoden zijn een welkome aanvulling, zoals meetinstrumenten over impact van open data en zoals aanvullende enquêtes die specifiek gericht zijn aan erfgoedinstellingen met open data. Het eenduidig definiëren van de term open data en een betrouwbare interpretatie van de term door instellingen in ondervragingen is mede bepalend voor het verzamelen van bruikbare vergelijkingsdata. Uit bestudering van GLAMetrics tools en uit de enquête komt naar voren dat het ontbreekt aan goede meetinstrumenten ten aanzien van het aanbod van open data, en het gebruik en bereik ervan. Het meten van toename en impact van open data door de tijd heen is problematisch. Daarom kunnen er op dit moment geen goede prognoses worden gedaan met het oog op de toekomst. Er kunnen hooguit verwachtingen uitgesproken worden door de erfgoedinstellingen zelf, maar dat geeft een eenzijdig beeld. Er is nog veel werk te doen door enerzijds computer- en informatiewetenschappers, maar ook door de erfgoedinstellingen die in hun organisatiedoelstellingen een prominente plek zouden moeten inruimen voor de ontwikkeling en inzet van open data en hun gebruikers meer zouden kunnen betrekken bij wat ze voor hen in huis hebben.

Literatuurlijst

- Baltussen, L., J. Oomen, M. Brinkerink, M. Zeinstra en N. Timmermans. "Open Culture Data: Opening GLAM Data Bottom-up." *Museums and the Web* 2013. Reds. N. Proctor & R. Cherry.
- Estermann, Beat. *Diffusion of Open Data and Crowdsourcing Among Heritage Institutions*. Conference Proceedings of the EGPA 2015 Conference, August 2015, Toulouse (France).
- Fritz, Catherine O., Peter E. Morris, en Jennifer J. Richler. "Effect Size Estimates: Current Use, Calculations, and Interpretation." *Journal of Experimental Psychology* 141.1 (2012): 2-18.
- Gurstein, Michael. "Open Data: Empowering the Empowered or Effective Data Use for Everyone?" *First Monday* 16.2 (2011): 1-5. 26 januari 2016. <<http://firstmonday.org/ojs/index.php/fm/article/view/3316/2764#p3>>.
- Nauta, Gerhard Jan en Wietske van den Heuvel. "Survey Report on Digitisation in European Cultural Heritage Institutions 2015". Enumerate, 2015.
- Terras, Melissa. "Opening Access to Collections: the Making and Using of Open Digitised Cultural Content." *Online Information Review* 39.5 (2015): 733-752.

Bijlagen

Dit onderzoek is onderbouwd met een aantal bijlagen. Deze zijn in aparte documenten verbonden aan dit eindverslag en kunnen per e-mail worden opgevraagd bij Tineke van der Meer (tineke.vandermeer@student.uva.nl). Het gaat om de volgende bijlagen:

Bijlage I: Instelling & Aanbieder - SPSS databestand met NL onderzoeksgegevens uit Estermann onderzoek

Bijlage II: Gebruik & Gebruikers - Survey Reacties Samengevoegd

Bijlage III: Gebruik & Gebruikers - OCD inventarisatie

Bijlage IV: GLAMetrics OCD instellingen

Bijlage V: Notulen bijeenkomsten