Gemeente Amsterdam Stadsarchief

noord-hollands archief

PROOF OF CONTEXTS

Een nieuwe toegankelijkheid voor archiefgebruikers

Eindrapport

Merel Geerlings (Stadsarchief Amsterdam) en Nico Vriend (Noord-Hollands Archief)

9 december 2024

Inhoud

Inleiding ... 3

Hoofdstuk 1. De ideeën ... 6

Het Archiveermonster helpen ... 6

Van boom naar netwerk ... 9

Hoofdstuk 2. Hoe dan? De uitvoering ... 13

In de casus: even met de hand ... 13

In de toekomst: geautomatiseerd ... 17

Hoofdstuk 3. Het ontwerp ... 18

De route ... 18

De prototypes ... 18

Hoofdstuk 4. De reacties ... 26

Tussenpresentatie Noord-Hollandse archivarissen ... 26

Tussenpresentatie Kring van Noord-Hollandse Archivarissen ... 26

Gebruiksonderzoek ... 26

Conclusie en vervolg ... 29

Vervolgstappen ... 29

Aandachtspunten ... 30

Met dank aan ... 32

Fotoverslag ... 33

Afbeelding titelblad:
The proof is in the pudding. Bij de marine worden hofmeesters opgeleid, verantwoordelijk voor de verzorging van opvarenden, 25 oktober 1958. Fotopersbureau De Boer, Noord-Hollands Archief

Inleiding

Stel, je hebt last van je kies. Je wilt het probleem snel verholpen hebben en je maakt een afspraak bij de tandarts. Wat verwacht je daar dan aan te treffen? Vast en zeker een geschoolde arts, een moderne behandelkamer en steriele instrumenten. Je ligt in de behandelstoel en de tandarts komt op je af met een roestige oude tang. “Komt goed,” zegt de tandarts, “ik doe dit altijd zo en het gaat altijd perfect.” Wat doe je dan? Waarschijnlijk sta je op, rent het gebouw uit, en denkt: hier kom ik nooit meer terug!

‘Kiezen trekken’, 26 februari 1963. Fotopersbureau De Boer, Noord-Hollands Archief

Gelukkig is er geen tandarts meer die nog zo te werk gaat. De tandarts gebruikt de nieuwste medische technieken en is geschoold om deze zo goed mogelijk toe te passen. Daar kun je van op aan.

Dit rapport gaat niet over de gebruikservaring bij de tandarts, maar bij het archief. Want hoe is deze als je op zoek bent naar informatie uit archieven? Ook hier hebben mensen een behoefte, namelijk om die bepaalde informatie te vinden. Ze gaan naar de juiste plek, een archief, en verwachten daar snel en efficiënt die gegevens te vinden. Maar wat er nu tevoorschijn komt is het equivalent van de roestige oude tang: de gebruiker wordt doorverwezen naar een online archiefinventaris, een voor velen ondoorgrondelijk fenomeen qua structuur en jargon. En weg is de nieuwe gebruiker, en die komt vast ook nooit meer terug…

Begrijp ons niet verkeerd: de inventaris als instrument om archief toegankelijk te maken heeft lange tijd goed gewerkt. Zoals vóór de moderne tandgeneeskunde iemand echt wel blij was dat die pijnlijke rotte kies eruit ging door middel van een tang, waren onderzoekers ook jarenlang tevreden

met de archiefinventaris. Je hoefde daardoor immers geen meters papier door te ploegen en je vond wellicht - uiteindelijk - wat je zocht. Maar de gedachte dat archief nu eenmaal 'moeilijk' is, en 'mensen maar hun best moeten doen om het te leren gebruiken', is inmiddels achterhaald. De archivaris kan niet alleen zijn relevantie in de maatschappij verhogen door zoveel mogelijk aan te sluiten bij hoe gebruikers het liefste willen zoeken. Maar methodologische en technologische ontwikkelingen om ons heen zorgen er ook voor dat het kán: de gebruiker een intuïtievere en gerichtere zoekervaring bieden.

De archiefsector heeft verschillende troeven in handen om de zoekervaring van gebruikers te verbeteren en optimaliseren, en aan te laten sluiten bij de algemene manier van zoeken op het wereldwijde web. Die troeven worden geboden door geavanceerde zoektechnologie, Linked Data en vormen van artificiële intelligentie zoals geautomatiseerde tekstherkenning (denk aan Handwritten Text Recognition, HTR) en taalmodellen (Large Language Model, LLM). Gebruikers zijn gewend aan Google, fulltext zoeken, willen filters toepassen en doorverwezen worden naar gerelateerde informatie, en zo snel mogelijk bij informatie uitkomen. Gebruikers willen niet gelaagd en door middel van veel doorklikken bij informatie uitkomen en willen niet op cursus om te weten wat er bedoeld wordt met 'Stukken betreffende bijzondere onderwerpen' of 'Aanvulling 2012'. Kortom, zij verwachten de gebruikelijke digitale zoekwijzen van de 21ste eeuw.

In het project *Proof of Contexts* (PoC) hebben we één van bovengenoemde troeven tot het verbeteren van de zoekervaring van gebruikers onderzocht: het daadwerkelijk omzetten van een hiërarchisch geordende archiefinventaris naar een netwerkstructuur. Het project is een samenwerking tussen het Noord-Hollands Archief (NHA) en het Stadsarchief Amsterdam (SAA). We hebben met ontwerpbureau Norday en datapartner Spinque een alternatief ontwerp gemaakt voor het zoeken door archiefinformatie. Samen met een groep archiefmedewerkers van Noord-Hollandse archiefinstellingen hebben we gekeken hoe we metadata anders kunnen modelleren om een nieuwe manier van zoeken en presenteren te faciliteren. Hierbij maken we gebruik van Records in Contexts, de nieuwe op linked data gebaseerde beschrijvingsstandaard voor archieven. Het ontwerp is vervolgens voorgelegd aan een gebruikersgroep.

Het doel van de PoC is om een alternatief voor de klassieke inventaris te ontwerpen en te testen, om aan te tonen dat het ook anders kan. Dit rapport is een inhoudelijk verslag van de afgelegde route, de denkstappen en keuzes die daarin gemaakt zijn. Om van PoC naar concreet plan te gaan moeten nog veel details verder onderzocht en bepaald worden. De PoC is dan ook bedoeld om het gesprek hierover op gang te brengen en als archiefsector dit gezamenlijk in de komende jaren verder uit te werken. We stimuleren iedereen om samenwerkingen aan te gaan, ideeën verder te ontwikkelen, om vervolgstappen en aanbevelingen op te pakken, verder uit te werken en daarover te publiceren.

In het rapport gaan we eerst in op de ideeën en de conceptuele uitwerking van de metadatamodellering. Hoofdstuk twee beschrijft hoe die modellering in de PoC in praktijk is gebracht en vervolgens geautomatiseerd tot stand kan komen, voor alle archieven. Het derde hoofdstuk handelt over het verloop van een zogenaamde ontwerpweek met Norday en Spinque, en toont de ontwerpen van de opgeleverde prototypes. De uitkomsten van een uitgevoerd gebruiksonderzoek staan in hoofdstuk vier, waarna een conclusie volgt en een deel over mogelijke vervolgstappen, aandachtspunten en aanbevelingen.

Veel leesplezier,

Merel Geerlings en Nico Vriend

Hoofdstuk 1. De ideeën

Het Archiveermonster helpen

Ken je het Archiveermonster? Dit monster komt voor in een van de vele verhalen die Andy en Terry beleven in het kinderboek *De waanzinnige boomhut*.[1] Op een dag komen de twee vrienden samen met hun vriendin Jill aan op een eiland waarop stapels archiefkasten staan. Ze gaan aan land en merken dat er weinig leven meer te vinden is: alles is gearchiveerd. Als ze een van de archieflades openen, komt het Archiveermonster tevoorschijn. Hij is woest, want de drie vrienden schoppen alles door de war! Het monster is dol op archiveren, en vond dit eiland dat nodig gearchiveerd moest worden. Hij besluit Andy, Terry en Jill óók te archiveren, "onder de I van IRRITANTE MENSEN DIE ZICH MET HUN EIGEN ZAKEN MOETEN BEMOEIEN! [...] Jullie zijn *irritante* mensen die zich overal mee bemoeien. Dus waar *anders* zou ik jullie moeten archiveren?", stelt hij. De vrienden komen vervolgens met allerlei andere opties: onder de eerste letters van hun namen bijvoorbeeld, of van hun beroep, of wellicht de kleur van hun ogen. "'STOP! STOP! STOP!', zegt het monster. 'Ik raak helemaal in war!' 'Sorry', zegt Jill. 'We proberen je gewoon te helpen.'"

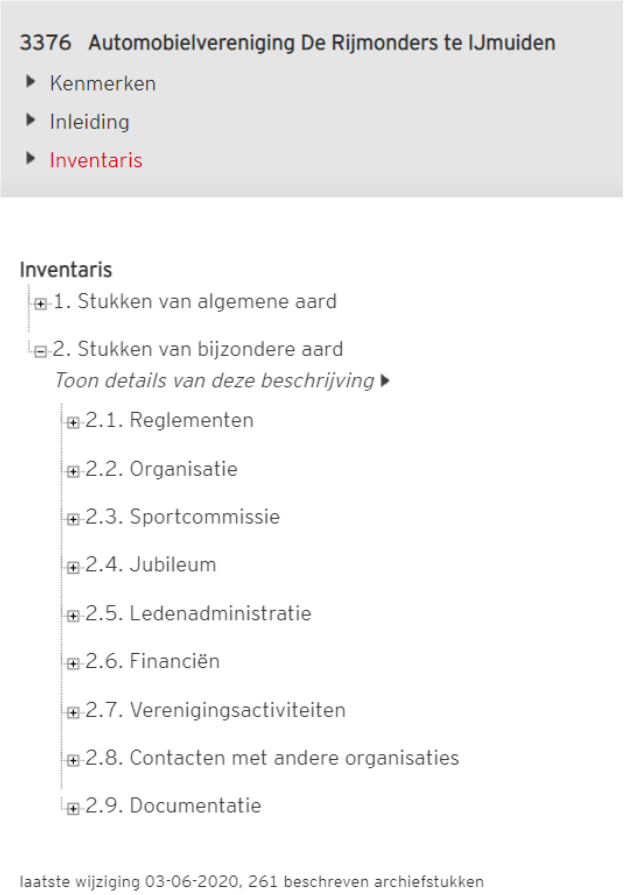
De auteurs van *De waanzinnige boomhut* tonen met het Archiveermonster - waarschijnlijk onbedoeld - een metafoor voor de dagelijkse archiveerpraktijk die al decennialang gaande is, in ieder geval in de eeuw na de publicatie van de wereldwijd invloedrijke *Handleiding voor het ordenen en beschrijven van archieven* van Muller, Feith en Fruin uit 1898. Deze praktijk is niet vreemd, in een analoge wereld wellicht zelfs de beste optie, maar heeft ook zo haar beperkingen. Want het Archiveermonster kan documenten maar in één lade stoppen, in één dossier, en onder één letter. In de digitale wereld hoeft dat niet meer: je kunt documenten in meerdere lades stoppen. Net als Terry, Andy en Jill moeten we het Archiefmonster helpen, want de archiefsector heeft de analoge wereld nog te veel laten doorwerken in de digitale.

Dat het bekende idee van één ding op één plek is gaan doorwerken in de digitale wereld is niet zo vreemd. Kijk eens naar Johan Leonard Lang, die op 1 juni 1899 door de Huidekooperstraat in Amsterdam rijdt in een auto. Hij trekt heel wat bekijks:

[1] Andy Griffiths en Terry Denton, *De waanzinnige boomhut van 117 verdiepingen* (Uitgeverij Lannoo, Tielt, 2020)

Johan Leonard Lang (1869-1952) op zijn eerste auto, een Franse Décauville, in de Amsterdamse Huidekoperstraat ter hoogte van de nummers 25-29, 1899. Jacob Olie, Stadsarchief Amsterdam

Die auto lijkt meer op een koets dan op een auto zoals we die nu kennen. Mensen kenden alleen koetsen, en wilden die mechaniseren. Dus leek de eerste auto op een koets. Hetzelfde zien we bij het archief, namelijk het oude idee van een document in één dossier op één plek in de archiefkast. In de klassieke inventaris staat een beschrijving van een document of dossier ook nog steeds op die ene plek in het geheel. En om dat stuk of dossier snel te kunnen vinden is er een ordeningsschema, een rubriekenstructuur, bedacht. Meestal wordt een nieuwe rubriek/map aangemaakt wanneer het aantal documenten te groot wordt en de lijst dan te lang. Het gevolg is vaak zeer gelaagde rubrieken/mappenstructuren.



Dat bestaat ook digitaal, denk aan de inventaris op de website, én aan een mappenstructuur in *Windows Verkenner*. Hoe vaak moet je wel niet in meerdere mappen kijken om de informatie die je zoekt te vinden of bij elkaar te krijgen? Hoe vaak gebeurt het niet dat een document over een onderwerp ook over andere onderwerpen gaat, en dat je dus op een heel andere plek moet zoeken? Als gebruiker klik je door zo'n hiërarchische boomstructuur heen en hoop je op een gegeven moment dat je alles gevonden hebt. Maar dat weet je nooit helemaal zeker.

Kijk eens naar het voorbeeld links, het archief van de Automobielvereniging De Rijmonders te IJmuiden. Daar is een dergelijke hiërarchische structuur gecreëerd. Digitaal kan op alle metadata worden gezocht.

Maar is metadata überhaupt nog wel van nut in een nieuwe werkelijkheid van records, waarbij de documenten *fulltext* doorzoekbaar zijn? Digitaal geboren archief is dit al, en HTR maakt het mogelijk om ook direct door de data van gedigitaliseerd analoog archief te zoeken. Je kan als het ware door het archief 'googlen,' een voor gebruikers fantastische ontwikkeling. Maar dat betekent niet dat metadata bij records in onbruik raken:

- Niet alle gegevens die je over een document wil weten staan ook in dat stuk. Een dagboekeigenaar schrijft bijvoorbeeld niet altijd zijn of haar naam in het dagboek. Metadata zal altijd van nut blijven als manier om informatie te zoeken.
- Metadata geeft ook context aan informatie: wie heeft het gevormd? Waarover gaat het, aan welke informatie is iets gerelateerd?
- Met *fulltext* zoeken is metadata een manier om het (potentiële grote) zoekresultaat te verfijnen, door middel van het gebruik van filters

Die metadata zit nu vast in die klassieke, hiërarchische archiefinventaris, zonder dat het verbonden is met andere gegevens. Het is dus zaak om die metadata zo te modelleren dat het bruikbaar is voor zoektechnieken die gebruikers online kennen en aan gewend zijn om te gebruiken.

De principes van linked data kunnen daarbij helpen, en Records in Contexts (RiC) als beschrijvingsstandaard. Daarvan is de basis dat informatie niet hiërarchisch wordt weergegeven, maar als netwerk van gegevens. Eén document of dossier kan dan wél verbonden zijn aan andere documenten en dossiers, en aan andere entiteiten zoals personen en locaties. Bij digitaal geboren informatieobjecten werken veel zaaksystemen en documentmanagementsystemen al op die manier.

De netwerkstructuur van RiC werkt daar veel beter dan de klassieke hiërarchische inventarisstructuur.

De op de koets lijkende auto heeft zich doorontwikkeld tot wat wij nu als auto herkennen. Met deze PoC hebben wij dat ook gedaan voor archiefinformatie.

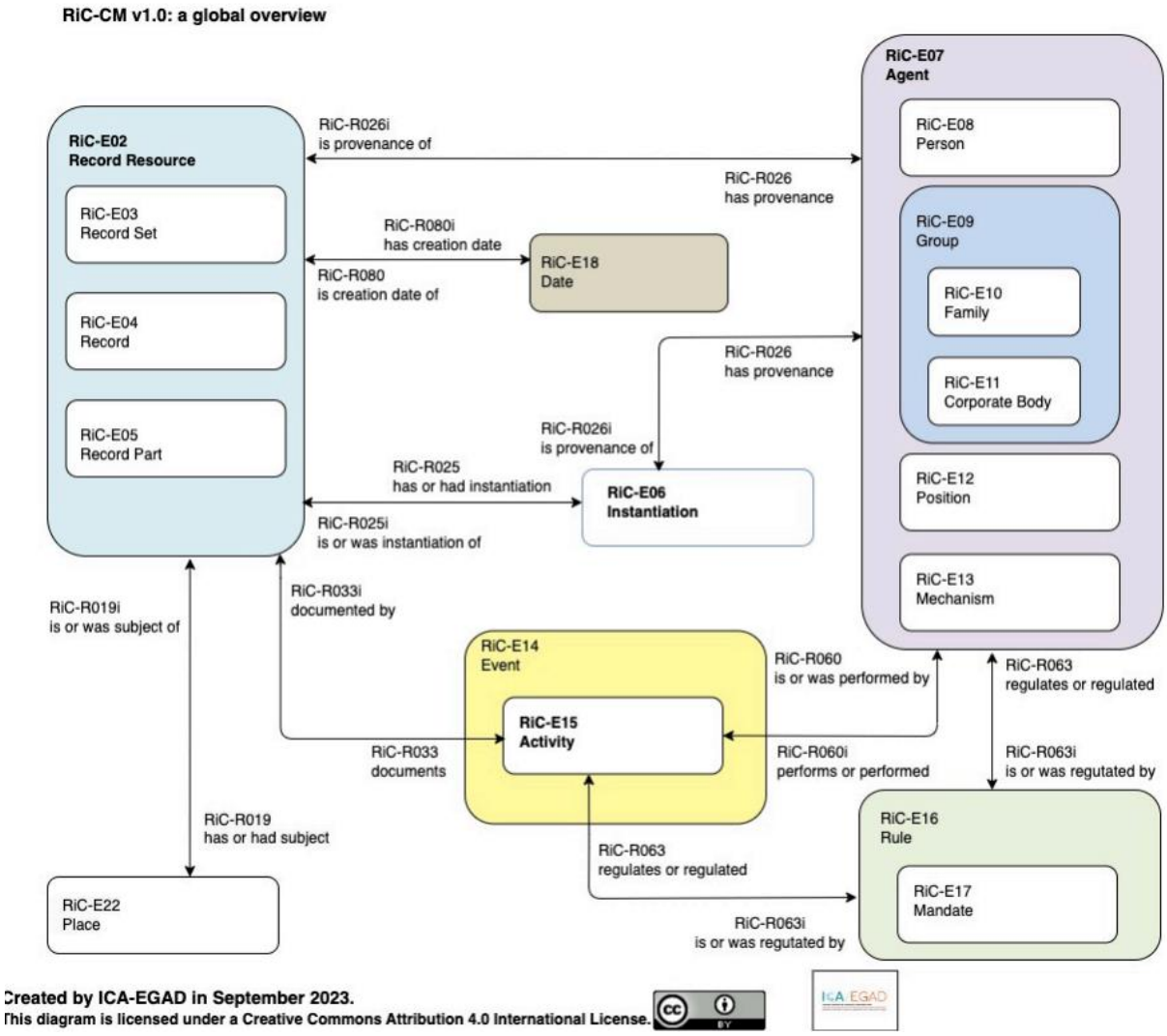
Van boom naar netwerk

Om te onderzoeken óf en hoe we een klassiek geordende archiefinventaris daadwerkelijk kunnen omzetten naar een netwerkstructuur zijn een flink aantal medewerkers van Noord-Hollandse archieven (Noord-Hollands Archief, Stadsarchief Amsterdam, Regionaal Archief Alkmaar, Archief Gooi en Vechtstreek, Westfries Archief en Gemeentearchief Zaanstad) bijeengekomen. Aan de hand van verschillende oefeningen is gekeken naar wat toegankelijkheid betekent voor archieven in de wereld van Linked Data.

In een van de oefeningen kropen de archivarissen in de huid van de gebruiker om te ontdekken wat voor hen de belangrijkste elementen zijn in het terugvinden van een specifiek record. Denkend vanuit het perspectief van de gebruiker werden vooral dataknooppunten als locatie-, persoons-, onderwerps- en tijdsaanduidingen genoemd, en pas daarna de archiefvormer. Natuurlijk is de archiefvormer een essentieel onderdeel om de context van records te duiden en begrijpen, maar moet die daarmee ook voor de gebruiker de voornaamste of zelfs enige zoekingang zijn? Op dit moment is dat wel de realiteit in de meeste van onze zoeksystemen.

Veel van de andere oefeningen gingen over het door archiefinstellingen gehanteerde ordeningsschema: de rubriekenstructuur. De conclusie was dat er met de sterk hiërarchisch ingerichte boomstructuur van klassieke inventarissen geen sprake is van een echt netwerk van gegevens, maar dat dat wel met terugwerkende kracht kan worden gemaakt. Zo zijn er gegevens die nu in rubrieken staan die beter tot hun recht komen op een andere plek in het netwerk. Deze omzetting - hetgeen centraal staat in de PoC - kan gerealiseerd worden door de toepassing van Linked Data als techniek, en Records in Contexts (RiC) als beschrijvingsstandaard.

Zoals de naam Records in Contexts al aangeeft, gaat het erom dat een stuk informatie (een record, of dat nu een document, afbeelding of audio-opname is) in context (wat dat ook maar mag zijn) wordt geplaatst. En contexts in meervoud, in alle contexten die bekend zijn. Om dat vast te leggen zijn in RiC de volgende entiteiten gedefinieerd:



Helemaal links is te zien dat je records (*Record Set*, *Record* en *Record Part*) kunt beschrijven, en die records kunt linken aan contextentiteiten als locatie (*Place*), datering (*Date*), gebeurtenissen (*Activity*) en allerlei actoren (*Person*, *Family* en *Corporate Body*). Alle entiteiten zijn onderling met elkaar verbonden door middel van relaties, en die relaties hebben ook een specifieke naam, zoals in het schema is te zien. Dat is de Linked Datastructuur die de basis van RiC vormt. Dit is nieuw voor archieven, want tot nu toe is elke inventaris een geïsoleerd fenomeen, de gegevens erin zijn niet of nauwelijks verbonden aan andere entiteiten.

RiC-E04	Record
Attribute ID	**Attribute Name**
RiC-A43	General Description
RiC-A22	Identifier
RiC-A28	Name
RiC-A03	Authenticity Note
RiC-A07	Classification
RiC-A08	Conditions of Access
RiC-A09	Conditions of Use
RiC-A10	Content Type
RiC-A21	History
RiC-A24	Integrity Note
RiC-A25	Language
RiC-A26	Legal Status
RiC-A35	Record Resource Extent
RiC-A38	Scope and content
RiC-A39	State
RiC-A40	Structure
RiC-A17	Documentary Form Type

Iedere entiteit binnen RiC heeft een uitgebreide lijst van gegevens (kenmerken, of *Properties* genaamd) die je over die entiteit kunt vastleggen. In de afbeelding links is te zien wat bijvoorbeeld over een record valt te zeggen, in maar liefst zeventien velden. Sommige kenden we al:

- *Identifier*: inventarisnummer
- *Name*: beschrijving
- *Record resource extent*: uiterlijke vorm

Maar er zijn er nog veel meer. Je kunt bijvoorbeeld een classificatie aangeven (personeel, financiën, etc.) of de redactionele vorm (*Documentary Form Type*) apart opslaan, in plaats van deze tekstueel in de beschrijving te noemen. Of denk aan de taal waarin het stuk geschreven is.

Met Records in Contexts in het achterhoofd heeft de groep archivarissen gekeken naar alle rubrieken van alle archieven van het Noord-Hollands Archief, dat zijn er zo'n 70.000. Die rubrieken blijken van allerlei aard te zijn, zie de afbeeldingen hieronder. De meeste van de gegevens uit de rubrieken blijken inderdaad beter tot hun recht te komen op een andere plek in het netwerk.

TOEGANGSCOD	CODE	VOLGNUMME	TITEL	DATERING	BEGINJAA	EINDJAA
141	2.1.03.	48	Jan van Styrum (1643-1680), gehuwd met Maria Steyn			
141	2.1.09.	74	Jan van Styrum (1721-1774), gehuwd met Anna Maria			
1614	2.1.1.2.2.	471	Jan van Sypesteyn C.A. I zoon (1670-1690)			
1614	2.1.1.2.2.	391	Jan van Sypesteyn en Cornelisz. (1633-1669), Maria v			
141	2.2.22.3.	670	Jan van Vollenhoven Willemszoon, gehuwd met Helen			
3934	6.2.	131	Jan van W (1759-1837)		1759	1837
3260	1.2.	18	Jan van Wickevoort Crommelin (1702-1740) en Elisab			
142	1.5.36.1.	6343	Jan Verloren			
1639	7.	925	Jan Voster	1992	1992	1992
1614	2.1.1.3.12	1608	Jan Willem Druyvesteyn François Constantijnsz. (175			
1614	2.1.1.3.12	1611	Jan Willem Druyvesteyn Jan Willemsz. (1786-1857) e			
936	5.5.	25	Jan Willem Reinier Tilanus (1823-1914)			
186.15	1.	5	Jan Willem Roeloffs Valk			
1604		50	Jan Willem van der Steur (1862-1945), A Xh			
1571	1.1.2.15.2	1019	Jannetje van Gilse 1838-1905 en Pieter Hendrik Testa			
3993	1.8.8.2.	10	Jansheren na 1625			
1604		5	Jansje van der Steur, C IX (1833-1903)			
3249	2.1.2.9.	232	Jansje van Zwieten			
1972	25.	2678	Janssen, E 1894-1896		1894	1896
1073	1.2.14.2.4	367	Jansstraat			
2106	13.2.2.2.1	5	Jansstraat			
2106	15.1.1.1.4	765	Jansstraat			
774	1.5.1.	2	Jansstraat 40			
18	2.2.6.2.	1935	Jansstraat 46, Prinsenhof, hulpgebouw Klein Heiliglan			
18	3.1.2.6.2.	7999	Jansstraat 46, Prinsenhof, hulpgebouw Klein Heiliglan			
5001	25.3.2.14.	17	Jansstraat e.o.			
3945		16	Jansstraat te Haarlem			
1518	3.3.2.	552	Jansstraat, Gemeentearchief			
1518	3.3.16.	1260	Jansstraat, Janskerk met Gemeentearchief			
5010	2.2.2.2.13	65	Jansweg/Parklaan			
1639	23.	974	Jantien de	1998	1998	1998
936	3.8.1.	5	Janus Red	1895	1895	1895
1627	2.6.1.	5	Japan			
2482		9682	Japan			
1627	2.6.	30	Japan, Zuid- en Oost-Azië			

TOEGANGSCOD	CODE	VOLGNUMMER	TITEL	DATERING	BEGINJAAR	EINDJAAR
615	7.1.	7504	Resumé of titles of Freudenthal's books, articles, lec			
615	7.1.1.	7508	A			
615	7.1.2.	7569	B			
615	7.1.3.	7631	C			
615	7.1.4.	7671	D			
615	7.1.5.	7718	E			
615	7.1.6.	7782	F			
615	7.1.7.	7811	G			
615	7.1.8.	7869	H			
615	7.1.9.	7926	I			
615	7.1.10.	7998	J			
615	7.1.11.	8003	K			
615	7.1.12.	8058	L			
615	7.1.13.	8128	M			
615	7.1.14.	8203	N			
615	7.1.15.	8252	O			
615	7.1.16.	8323	P			
615	7.1.17.	8373	Q			
615	7.1.18.	8379	R			
615	7.1.19.	8468	S			
615	7.1.20.	8548	T			
615	7.1.21.	8627	U			
615	7.1.22.	8651	V			
615	7.1.23.	8703	W			
615	7.1.24.	8818	Y			
615	7.1.25.	8822	Z			
616	1.	64	Stukken van algemene aard			
616	1.1.	65	Jaarverslagen			
616	2.	79	Stukken van bijzondere aard			
616	2.1.	80	Organisatie en personeel			
616	2.1.1.	81	Organisatie			
616	2.1.2.	112	Financiën			

De gegevens uit de rubrieken blijken grofweg gecategoriseerd te kunnen worden naar soorten acties:

- Gegevens van een rubriek toevoegen als een kenmerk bij het record:
  - *Classificatie*, zoals personeel, financiën
  - *Documentary form type*, zoals jaarverslag, notulen
  - *Date*, zoals jaartal of datum

- Gegevens van een rubriek toevoegen als een andere entiteit:
    - Persoon, alle namen van personen horen in de entiteit *Person*, en moet daar omschreven worden. Met een relatie kan worden aangegeven of het om bijvoorbeeld een archiefvormer of een vervaardiger gaat
    - Organisatie, idem als personen
    - Locatie
    - Onderwerpen

Ook zijn er rubrieksnamen die grotendeels betekenisloos zijn voor een gebruiker. Omdat niemand op die woorden zoekt, en ze ook geen context geven aan een stuk. Die gegevens hebben we in de PoC genegeerd:

- Stukken van algemene aard, stukken van bijzondere aard
- Overig, varia, diversen
- Documentatie
- Stukken waarvan het verband met het archief niet duidelijk is
- Aanvulling
- Letters (A, B, C, etc.)

Tot slot zijn er rubrieksnamen die voor meerdere interpretaties vatbaar zijn. Bijvoorbeeld Concertgebouw: dat is een gebouw, een locatie en een organisatie. Wat wordt dan precies met de rubriek bedoeld? Alleen een verdere blik op de inventaris, of de onderliggende stukken zelf, kan dan duidelijkheid bieden.

Als de bewerkingsacties worden uitgevoerd, wordt de rubriekstructuur omgezet zonder betekenis te verliezen en blijft er een platte lijst van records over. De records zijn dan op zichzelf beter beschreven en bovendien op de juiste manier verbonden aan contextentiteiten. Door de omzetting blijft niet alleen de bestaande betekenis behouden, maar bieden we de gebruiker juist méér betekenis en contexten aan. Het voegt iets toe, in plaats van dat het iets wegneemt. In hoofdstuk drie van dit rapport, waar het ontwerp aan bod komt, wordt duidelijk waarom het zo belangrijk is om metadata op deze manier te modelleren. Eerst gaan we in het volgende hoofdstuk in op de uitvoering van de omzetting van de metadata.

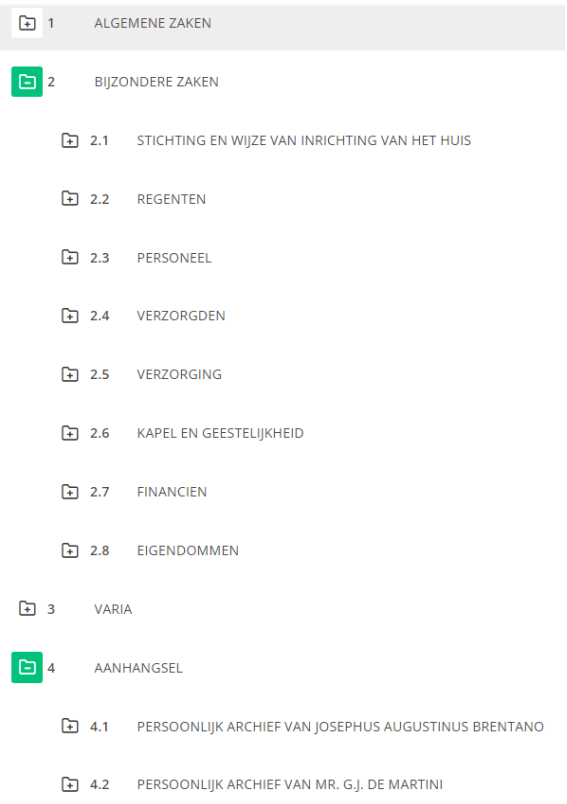
Hoofdstuk 2. Hoe dan? De uitvoering

In hoofdstuk 1 hebben we conceptueel gekeken naar het omvormen van de klassieke inventaris naar een netwerk van gegevens. Hoewel de groep archivarissen die daaraan werkte erg enthousiast was over de gezamenlijk gezette denkstappen, bleef echter één belangrijke vraag overeind staan: 'hoe dan?' In dit hoofdstuk staat beschreven hoe we eerst een casusarchief met de hand hebben bewerkt, en daarna wat we nu weten over technieken om dat geautomatiseerd te kunnen doen.

In de casus: even met de hand

Om te testen of de ideeën ook in de praktijk werken, hebben we één archief uitgezocht om als casus te gebruiken. Dat moest aan de volgende criteria voldoen:

- Het moest verschillende ordeningsschema's bevatten, zoals stukken van algemene aard, bijzondere onderwerpen, en gedeponeerd archief.
- Het mocht niet te groot of te klein zijn qua aantal records
- Het moest gescande records hebben



We zijn uitgekomen bij het archief met toegangsnummer 439 van het Stadsarchief Amsterdam: Archief van de Regenten van de Stichting Rooms-Katholiek Burger Oude-mannenhuis. Een archief dat onder andere over nu nog actuele maatschappelijke thema's als ouderenzorg en -huisvesting gaat. Op de afbeelding hiernaast is de hoofdstructuur van het archief te zien. We hebben deze structuur geanalyseerd en aan de hand van de ideeën uit de conceptuele fase handmatig bewerkt. Dit hebben we gedaan in een testomgeving binnen het collectiebeheersysteem van het Stadsarchief Amsterdam, waar we de records (de bestanddelen van het archief) en het archiefblok naartoe hebben gekopieerd. Zo konden we ongestoord aan de slag.

Stap 1: betekenisloze rubrieken als tussenlaag verwijderen

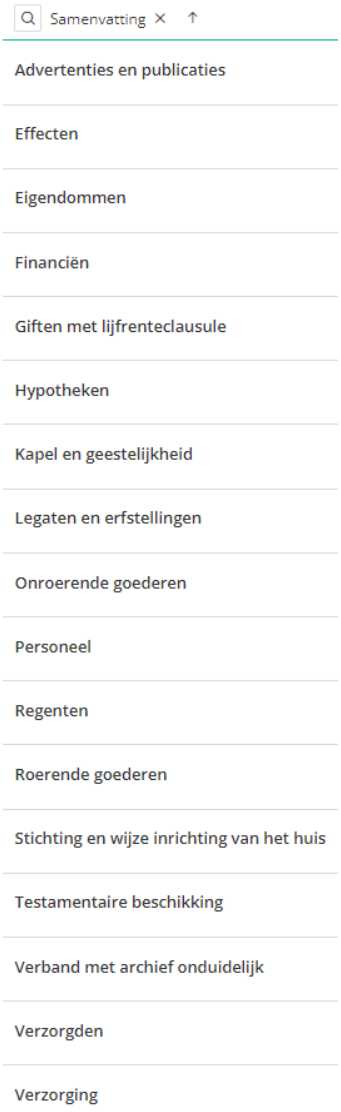
In dit archief is goed zichtbaar dat er rubrieken zijn die betekenisloos zijn voor een gebruiker. In dit archief gaat het om de rubrieken:

- Algemene zaken
- Bijzondere zaken
- Varia
- Aanhangsel

Deze rubrieken hebben we voor de praktische uitvoering van de PoC verwijderd.

Stap 2: gegevens uit rubrieken die categorieën zijn toevoegen aan het record

Binnen de rubriekenstructuur van dit archief worden veel termen gebruikt die zich goed lenen om een classificatieschema voor aan te maken. Denk aan 'Personeel' en 'Financiën'. Gedacht in RiC is dit een kenmerk/property bij het record (rico:Classification). We hebben daarvoor een nieuw veld aangemaakt en de termen van de rubrieken in de vorm van een keuzelijst daaraan 'as-is' gehangen (zie de afbeelding rechts).

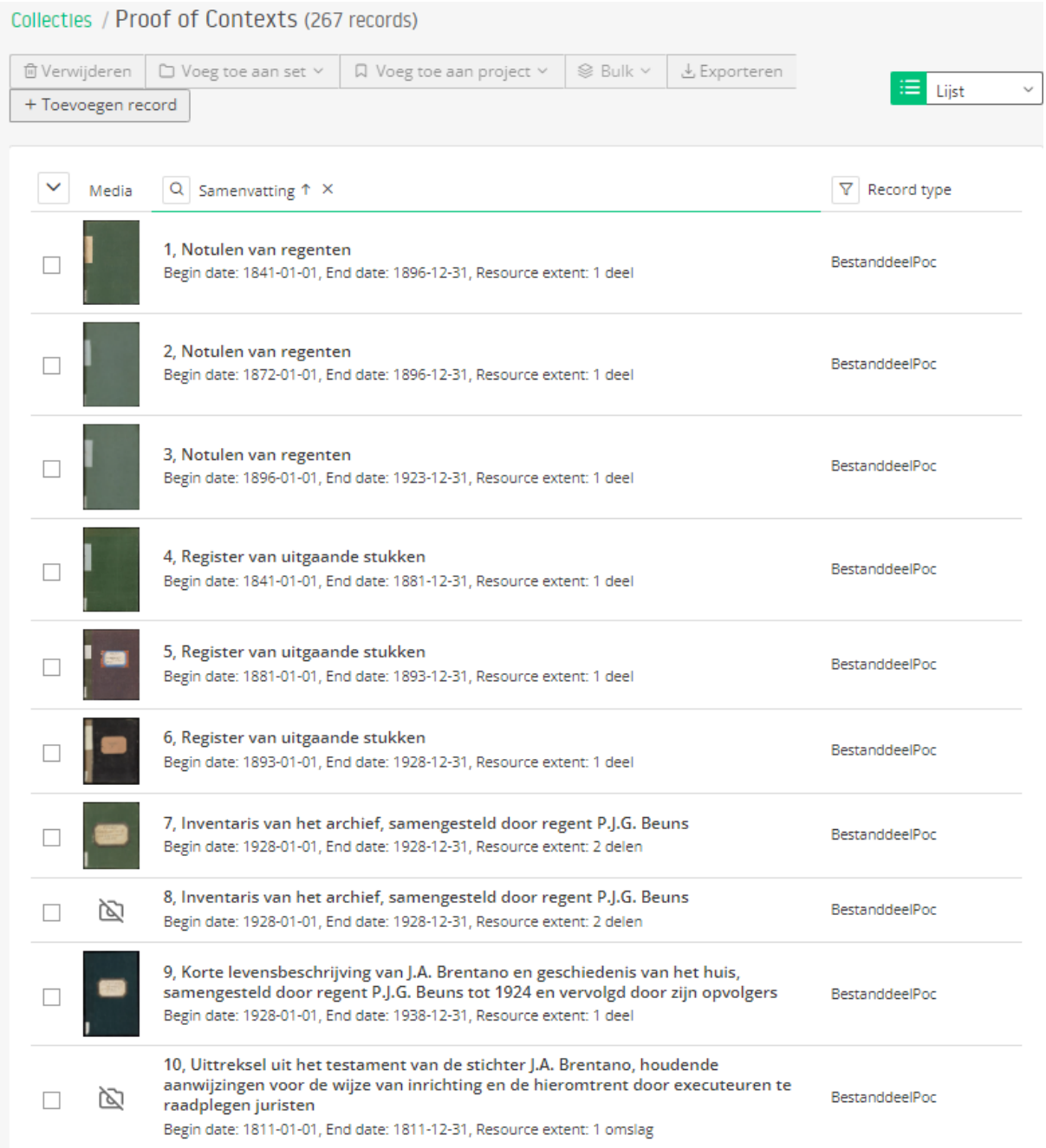


Deze keuzelijst is de basis van de filtermogelijkheden die in het ontwerp van de PoC worden aangeboden, zoals uit hoofdstuk drie zal blijken. Deze is echter niet zonder beperkingen, zoals uit de afbeelding rechts blijkt, want de lijst is niet van hoge kwaliteit. Termen als 'eigendommen' en 'effecten' hebben natuurlijk te maken met financiën en zouden idealiter aan een dergelijke term moeten worden gelinkt via een terminologiebron. Op die manier wordt het zoeken doeltreffender en verzeilt een gebruiker niet in een woud van duizenden verschillende termen. Maar op dit moment ontbreekt een goede terminologiebron voor classificatie/categorie. Het ontwikkelen hiervan is een voorwaarde om geautomatiseerd aan de slag te kunnen gaan met alle archieven. Omwille van de oefening hebben we de termen uit de rubrieken nu overgenomen zoals ze zijn.

Stap 3: archiefvormers identificeren en toevoegen aan entiteit Persoon

In dit archief zijn er twee rubrieken met 'aanhangsels', bestaande uit archiefmateriaal dat gevormd is door anderen (zie afbeelding linksonder). Het gaat hier om twee personen, J.A. Brentano en G.J. de Martini, waarvoor we de entiteit *Persoon* hebben. Alle records die onder die rubriek hingen, hebben we gerelateerd aan de *Persoon*, met de rol van *Archiefvormer* daarbij. De rubrieken zijn vervolgens voor de praktische uitvoering van de PoC verwijderd.

Het resultaat van deze drie stappen is dat alle gegevens uit de rubrieken ofwel toegevoegd zijn als een kenmerk bij het record, ofwel als een andere entiteit met een relatie daarnaartoe. Het resultaat is een platte lijst van records:



We besloten echter om nog een stap verder te gaan in de PoC en ook op het niveau van de archiefbeschrijvingen verrijkingen toe te passen.

Stap 4: informatie uit de beschrijvingen van de records kopiëren en toevoegen

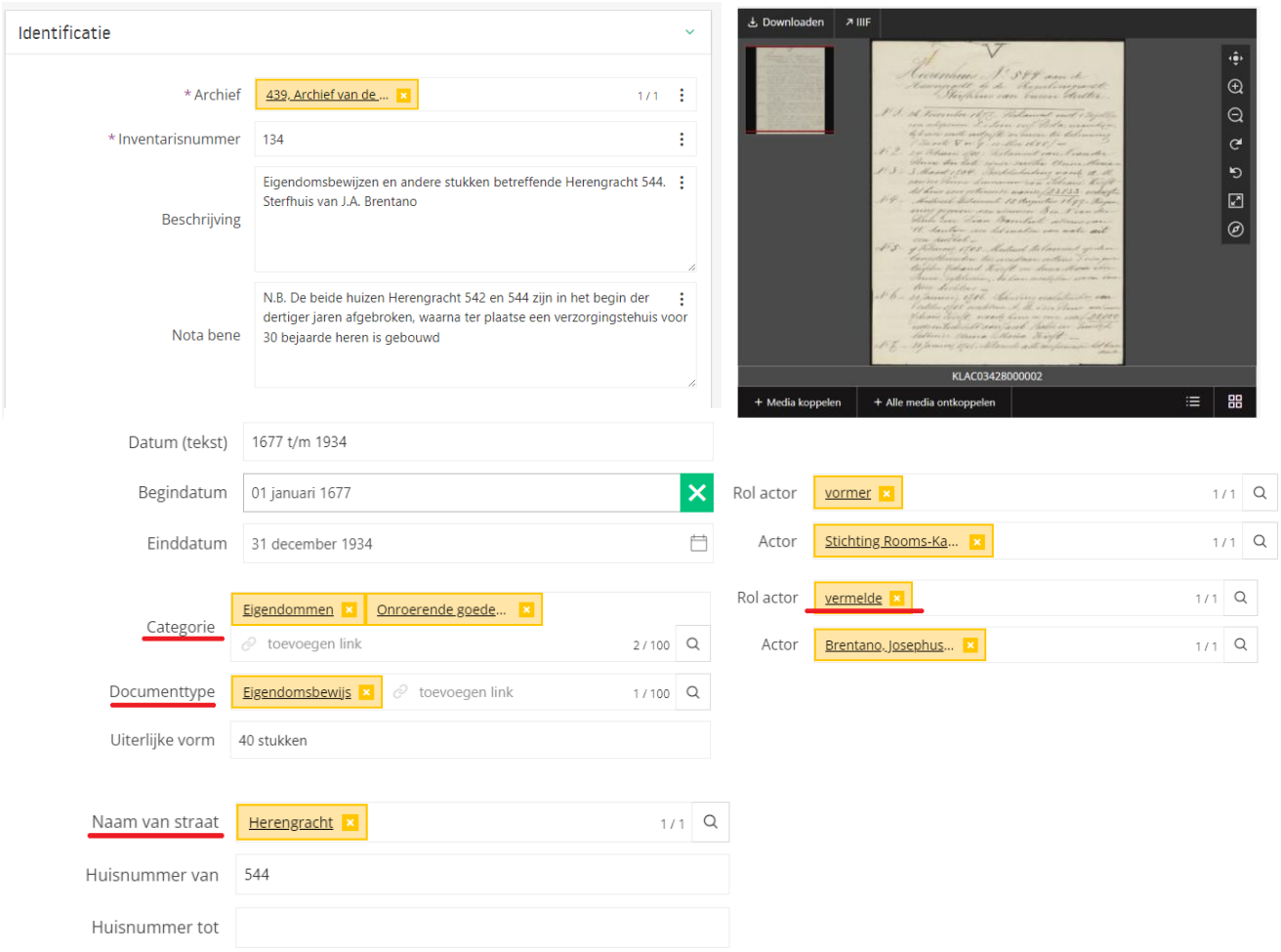
Het casusarchief bevat behoorlijk uitgebreide archiefbeschrijvingen. Daar kan eenvoudig een standaardisatieslag mee worden geslagen:

- De verzamelbeschrijvingen (vb's) kunnen worden opgeheven door de gegevens van de verzamelbeschrijving te verhuizen naar de beschrijving van het record.
- In de beschrijvingen van de records staan gegevens die volgens RiC beter op een andere plek kunnen staan:
    - Veel beschrijvingen beginnen met een redactionele vorm, zoals 'brief'. In RiC is dat een apart kenmerk bij record. Ook dit gegeven is te standaardiseren door gebruik te

maken van een nog te ontwikkelen terminologiebron voor documenttypen. Wij hebben een extra veld bij het record aangemaakt, de keuzelijst gemaakt en de gegevens aan het record gekoppeld. De lijst is nu een platte lijst met termen, maar is idealiter gekoppeld aan gestructureerde termen, met horizontale en verticale (hiërarchische) relaties, zoals in een SKOS (Simple Knowledge Organization System) datamodel.

- Vaak hebben de beschrijvingen een adres in Amsterdam. Ook dat gegeven hebben we opgenomen als kenmerk bij het record. Locaties buiten Amsterdam zijn net zo relevant, maar zijn voor de praktische uitvoering van de PoC even genegeerd.
- De beschrijvingen bevatten veel persoons- en organisatievermeldingen. We hebben die betreffende personen en organisaties als item aangemaakt, en verbonden aan het record met de relatie ‘vermelde’.

De afbeeldingen hieronder laten zien welke gegevens door de verschillende bewerkingsacties nu direct bij het record staan, zie de kenmerken met de rode streep eronder.



In de toekomst: geautomatiseerd

Kunnen de acties die we nu met de hand hebben gedaan in de toekomst ook geautomatiseerd worden? Om die vraag te beantwoorden is een groep technisch experts ingeschakeld: Rick Companje, Mark Lindeman, Antoinet Nijssen, Niek Verhoeff en Ivo Zandhuis. Uit een brainstormsessie kwam naar boven dat een vrij eenduidig - en voor de experts enigszins voordehandliggend - stappenplan kan worden voorgeschreven:

1. Je hebt je datamodel klaar en je terminologiebronnen op orde. In het geval van classificatie/categorie en documentsoort moet nog een terminologiebron worden ontwikkeld. Dit geldt ook voor actoren (personen, families, organisaties en organisatieonderdelen).
2. Je exporteert de metadata van het record, waarbij je per record de padnaam (archief→rubriek→beschrijving) van de rubriekenstructuur meeneemt. Je slaat dit op.
3. Je past keywordextractie toe op de geëxporteerde gegevens door de hele padnamen van de records door een taalmodel te halen. Je vraagt het model om een lijst van de voornaamste termen te genereren. Die lijst is niet alleen de basis voor de filtermogelijkheden, het is ook de basis voor verrijkingen die je kunt toepassen.
4. Je verrijkt de keywords door ze te koppelen aan:
    a. Terminologiebronnen, zoals bij classificatie/categorie en documentsoort
    b. Andere entiteiten (zoals Personen, Archiefvormers en Organisatieonderdelen) in je datamodel. Als de entiteit nog niet bestaat, kun je die nieuw aanmaken en koppelen.

Technieken die voor deze stappen nodig zijn, zijn Named Entity Recognition en Named Entity Linking. Open programma's als Open Refine zijn daarbij behulpzaam. En het is handig om de programmeertaal Python en de querytaal SPARQL te beheersen.

De expertgroep gaf aan dat er stapsgewijs gewerkt kan worden: entiteit voor entiteit, verrijken en standaardiseren waar je kan. Je begint bijvoorbeeld met een makkelijke entiteit, zoals Personen, omdat die relatief eenvoudig te identificeren zijn. En daarna kan je bijvoorbeeld aan de slag met documentsoorten. De klassieke archiefinventaris zoals we deze nu nog kennen, hoeft dan ook niet meteen radicaal overboord gegooid te worden. In de PoC hebben we dat omwille van de gedachteoefening wel gedaan. Het beter gebruikmaken van het netwerk door gegevens toe te voegen is complementair aan de boomstructuur van de klassieke archiefinventaris. Het op een andere en verrijkte manier aanbieden van de classificaties of categorieën bestaat dan náást het aanbieden van een gelaagde rubriekenstructuur. De archiefboom blijft één van de zoekopties voor gebruikers, maar we creëren daarnaast toegangen op andere entiteiten en kenmerken in het netwerk. Onze verwachting is wel dat de boom door gebruikers steeds minder als zoekingang gebruikt gaat worden, naarmate we steeds meer kunnen verrijken. De boom zal daarmee wellicht niet verdwijnen, maar wel steeds minder relevant worden.

Hoofdstuk 3. Het ontwerp

Het op conceptueel niveau uitdenken van nieuwe zoekingangen en datamodellering is één ding, het daadwerkelijk maken van een nieuwe zoekinterface voor gebruikers is een tweede. Voor deze fase is de hulp ingeschakeld van ontwerpbureau Norday en technisch partner Spinque en zijn we opnieuw in de huid van de gebruiker gekropen, ditmaal om te ontdekken hoe de gebruiker zou willen zoeken en resultaten gepresenteerd zou willen krijgen.

De route

Om samen met Norday en Spinque tot een waardevol ontwerp te kunnen komen, zijn we gestart met de probleembepaling en afbakening van de PoC. Hiervoor zijn verschillende zogenaamde experttalks gehouden waarin medewerkers van het NHA en het SAA vertelden hoe archieven in elkaar steken en wat er op het gebied van toegankelijkheid zoal te verbeteren is. De focus van de PoC ligt op het presenteren van records op een andere, nieuwe manier. Kunnen we daar een aantrekkelijk en zelfverklarend ontwerp voor maken?

We zijn met elkaar gaan schetsen. Hoe zou een archiefpagina er uitzien (dus van het archiefblok, het fonds, waar je nu veelal de inleiding en de inventaris gepresenteerd krijgt)? En hoe ziet de detailpagina van een record eruit?

Alle ontwerpen zijn doorgenomen, en de deelnemers hebben gestemd op de voorstellen die ze het meest aanspraken. Die uitkomst was de basis van het visuele prototype.

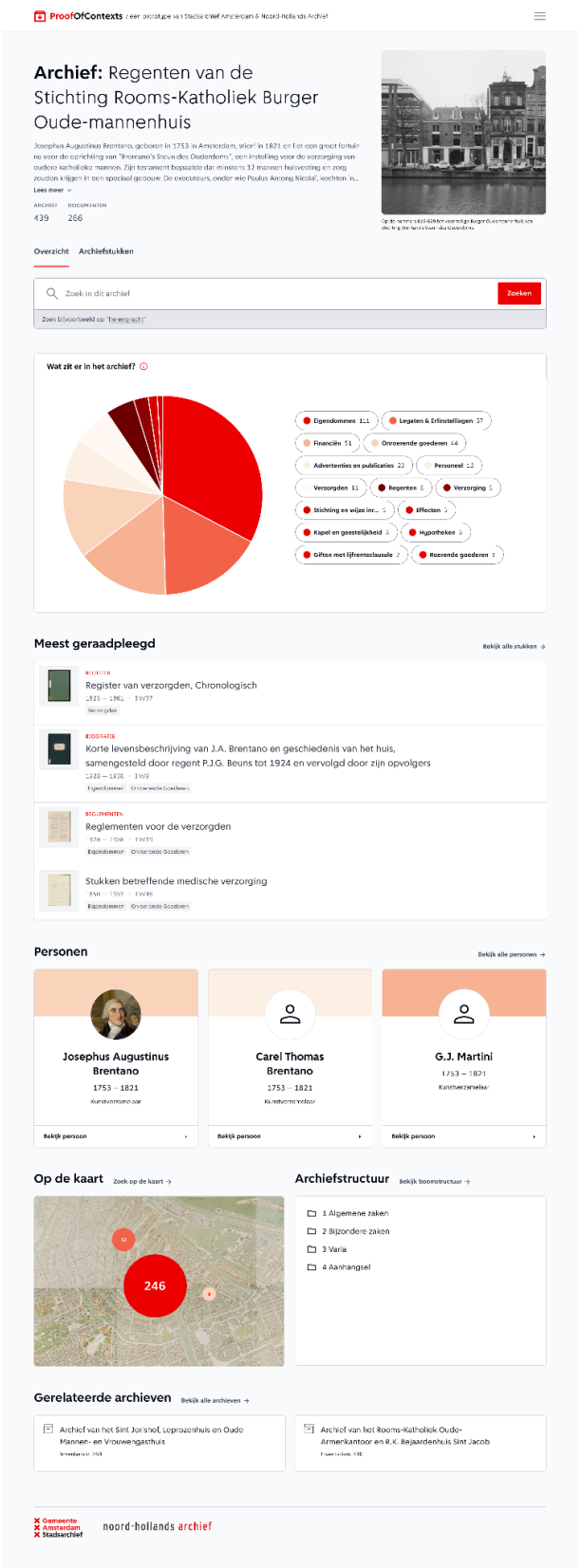
Het ontwerp gaat uit van de presentatie van één specifiek archief, het in het vorige hoofdstuk genoemde casusarchief. De zoekroute náár deze specifieke archieftoegang is buiten beschouwing gelaten. De met de hand bewerkte dataset is ingestuurd naar Spinque, die hem heeft doorgrond, en heeft onderzocht wat de mogelijkheden ermee zijn.

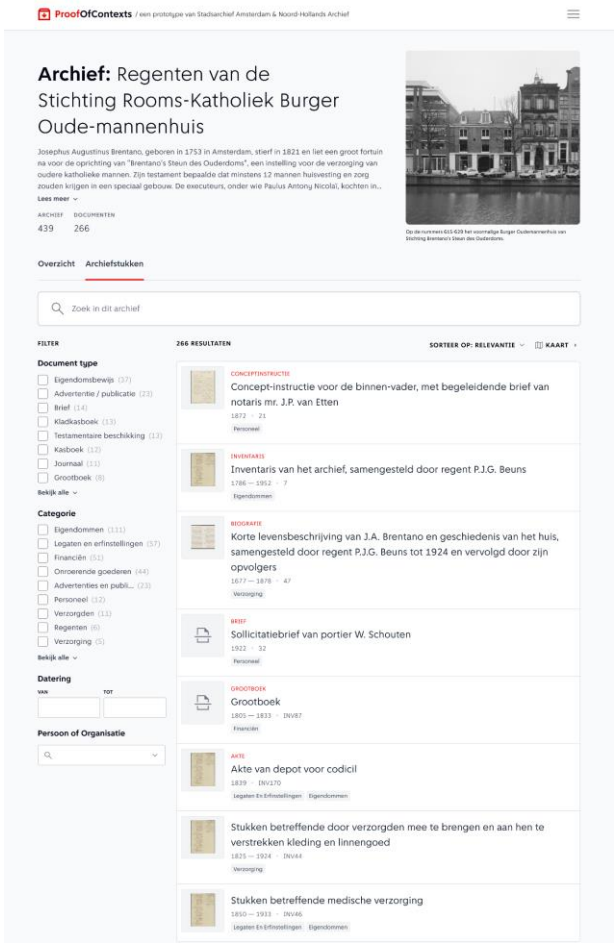
Een wand vol schetsen, met gele stipjes om te stemmen op favoriete onderdelen

De prototypes

Er zijn twee prototypes gemaakt: een visueel en een technisch ontwerp. Het visuele design is op deze plek te bekijken: SAA-NHA Design Prototype - Design. Het is gebaseerd op de data die we hebben gemaakt voor de PoC, maar is geen werkende website. Dat wil zeggen dat de data daarin niet live is te bevragen. Daarom heeft Spinque ook een technisch prototype gemaakt. Zo kunnen we beter testen of de aanpassingen in de metadata ook leiden tot een betere zoekervaring.
Het visuele prototype bestaat uit ontwerpen voor een archief (archiefblok), een record (uitgebreid) en een record (eenvoudig).

Het ontwerp van het gehele archief begint bovenaan met de titel en de inleiding van het archiefblok. Daaronder is een zoekbalk te zien, gevolgd door twee tabbladen: ‘Overzicht’ en ‘Archiefstukken’:





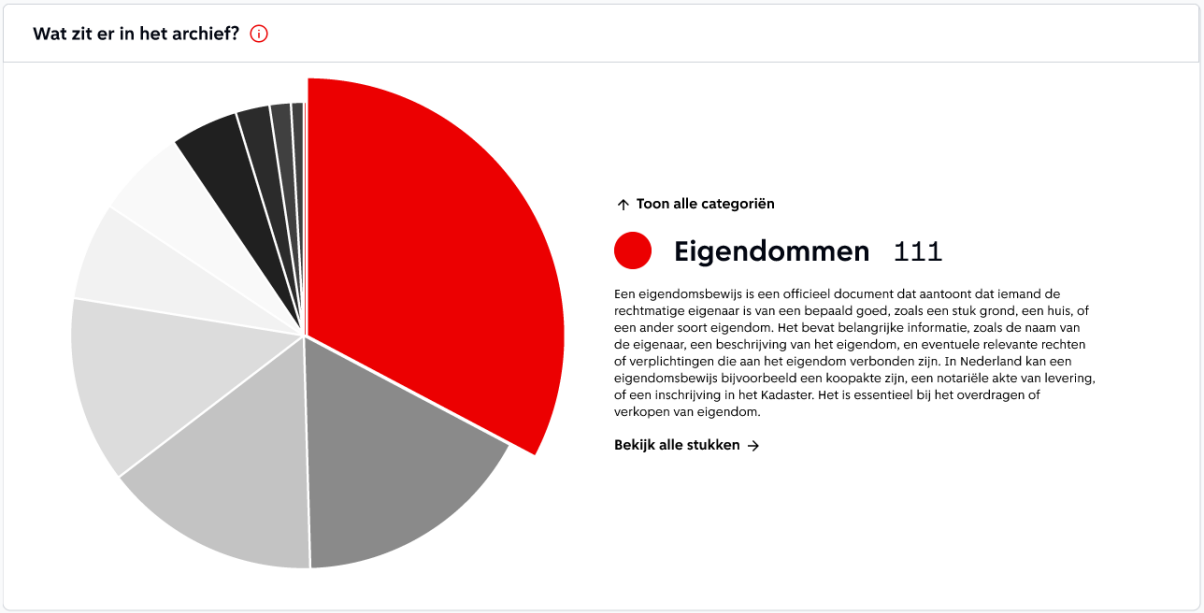
Afbeelding links: tabblad ‘Overzicht’.
Afbeelding rechts: tabblad ‘Archiefstukken’

Het tabblad ‘Overzicht’ begint met een *dashboard*, met daarin informatie over wat er zoal in dit archief te vinden is. Het is bedoeld voor de gebruiker die zich wil oriënteren. Het dashboard bevat de volgende onderdelen:

- *Wat zit er in het archief?* Hier staat een taartdiagram, gevuld met de categorieën die aan de records van dit archief gekoppeld zijn. Zo zie je in één oogopslag de thema’s waar de stukken zoal over gaan. Het is bedoeld om de gebruiker een gevoel van overzicht te geven.
- *Meest geraadpleegd.* Van veel archieven weten we heel goed welke records mensen het liefste zien. Dit is voor veel mensen dus een handige shortcut naar een record, of het geeft een idee van wat er zoal in het archief te vinden is.
- *Personen.* Archieven gaan vaak over mensen. We kunnen laten zien welke personen gelinkt zijn aan een specifieke archieftoegang. Je kunt ervoor kiezen om via de persoon verder te zoeken naar records.
- *Op de kaart.* Sommige archieven hebben een sterke geografische component. Dan kan het handig zijn om op de kaart te zoeken.
- *Archiefstructuur.* Zoals aangegeven in hoofdstuk 2 kan de klassieke inventaris nog steeds worden aangeboden als een manier van zoeken, naast andere vormen van weergave. Dus voor wie nog even in de inventaris wil kijken, kan hier terecht.
- *Gerelateerde archieven.* Hier kun je inzichtelijk maken dat er gerelateerde archieven zijn en de gebruiker een hint geven om daar ook een kijkje te nemen. Er zijn bijvoorbeeld archiefvormers die verschillende archiefblokken hebben.

De gegevens uit het dashboard worden zo veel mogelijk geautomatiseerd en datagedreven gegenereerd. De meest geraadpleegde stukken kunnen bijvoorbeeld voortkomen uit statistische gegevens en ook het taartdiagram kan worden opgebouwd uit cijfers over de hoeveelheid inventarisnummers (of eventueel scans) per categorie. Deze datagedreven aanpak heeft als belangrijk voordeel dat het praktisch ook uitvoerbaar is. Het is immers niet te doen om dergelijke overzichten handmatig te maken voor duizenden verschillende archieftoegangen, al kan er altijd een mogelijkheid worden geboden om handmatig aanpassingen te doen.

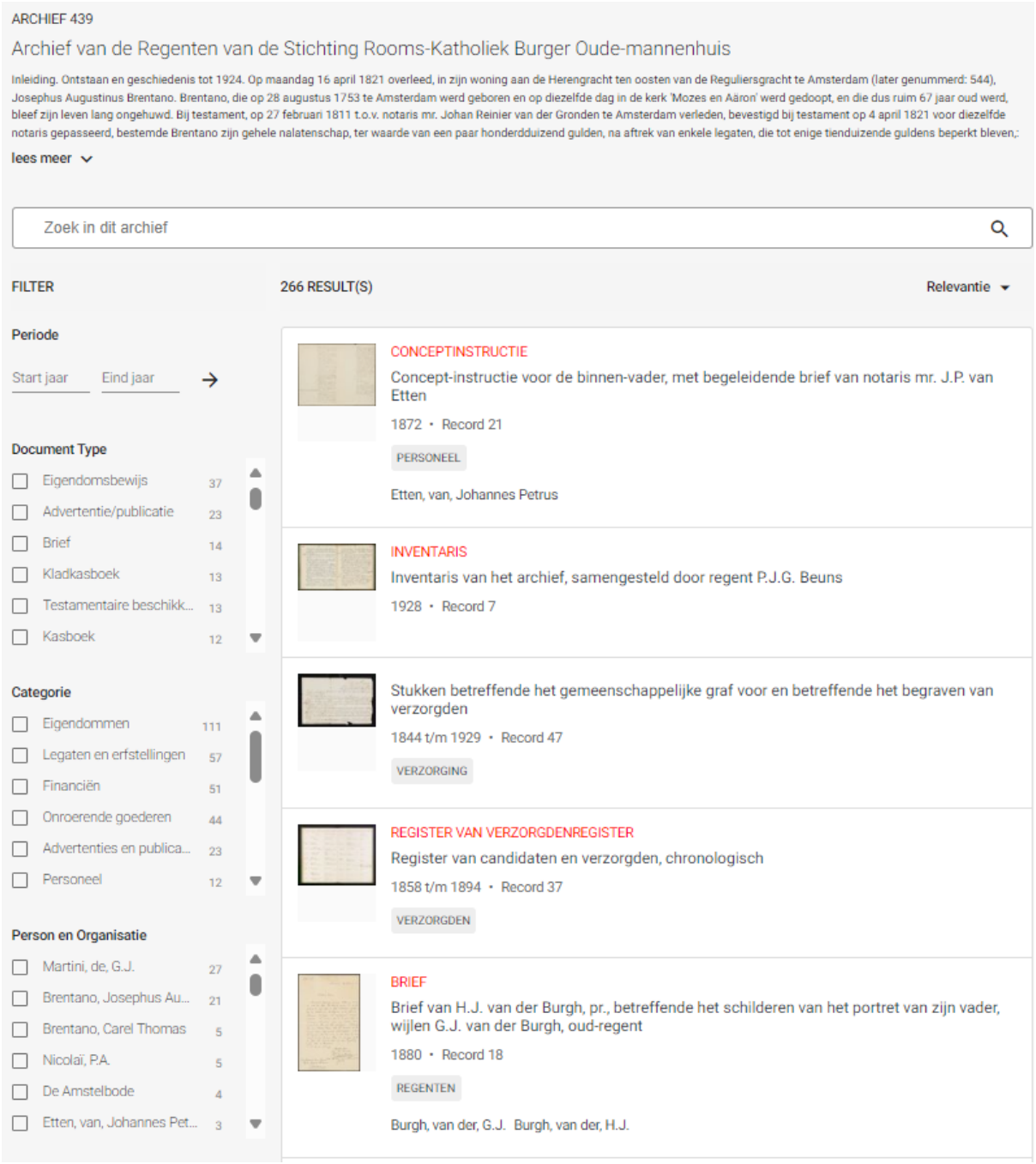
De datagedreven aanpak heeft nog meer voordelen, voortkomend uit de principes van linked data. De databronnen waarmee gekoppeld wordt, zoals een terminologiebron voor classificatie/categorie of persoonsreconstructies, bevatten namelijk allerlei verrijkte gegevens die ook hier getoond kunnen worden. In ons voorbeeld zie je bijvoorbeeld geboorte- en sterfjaren en beroepen terug bij de getoonde personen - gegevens die uit Wikidata afkomstig zijn - of een uitlegtekst als je klikt op één van de categorieën uit de taartdiagram - gegevens die afkomstig zouden zijn uit de (nog te maken) terminologiebron. Een van de grote voordelen hiervan is dat je gegevens maar op één plek hoeft op te slaan, namelijk in de terminologiebron, om deze op meerdere plekken te tonen, hergebruiken en van waarde te laten zijn.



Taartdiagram na het klikken op één van de categorieën

De datagedreven insteek heeft echter ook een keerzijde, want het dashboard kan alleen worden opgebouwd op basis van nu beschikbare metadata. Het kan daarmee een scheef beeld geven van de werkelijkheid, bijvoorbeeld doordat minder gerepresenteerde groepen of niet geïndiceerde personen ontbreken in dit soort overzichten. Om de gebruiker zich hiervan meer bewust te maken, kan bij ieder gewenst onderdeel uitleg worden gegeven over de totstandkoming van de gegevens. In het ontwerp zie je dit terug in de vorm van een ‘i’ (met *hoover-over*) achter het onderdeel *Wat zit er in het archief?*.

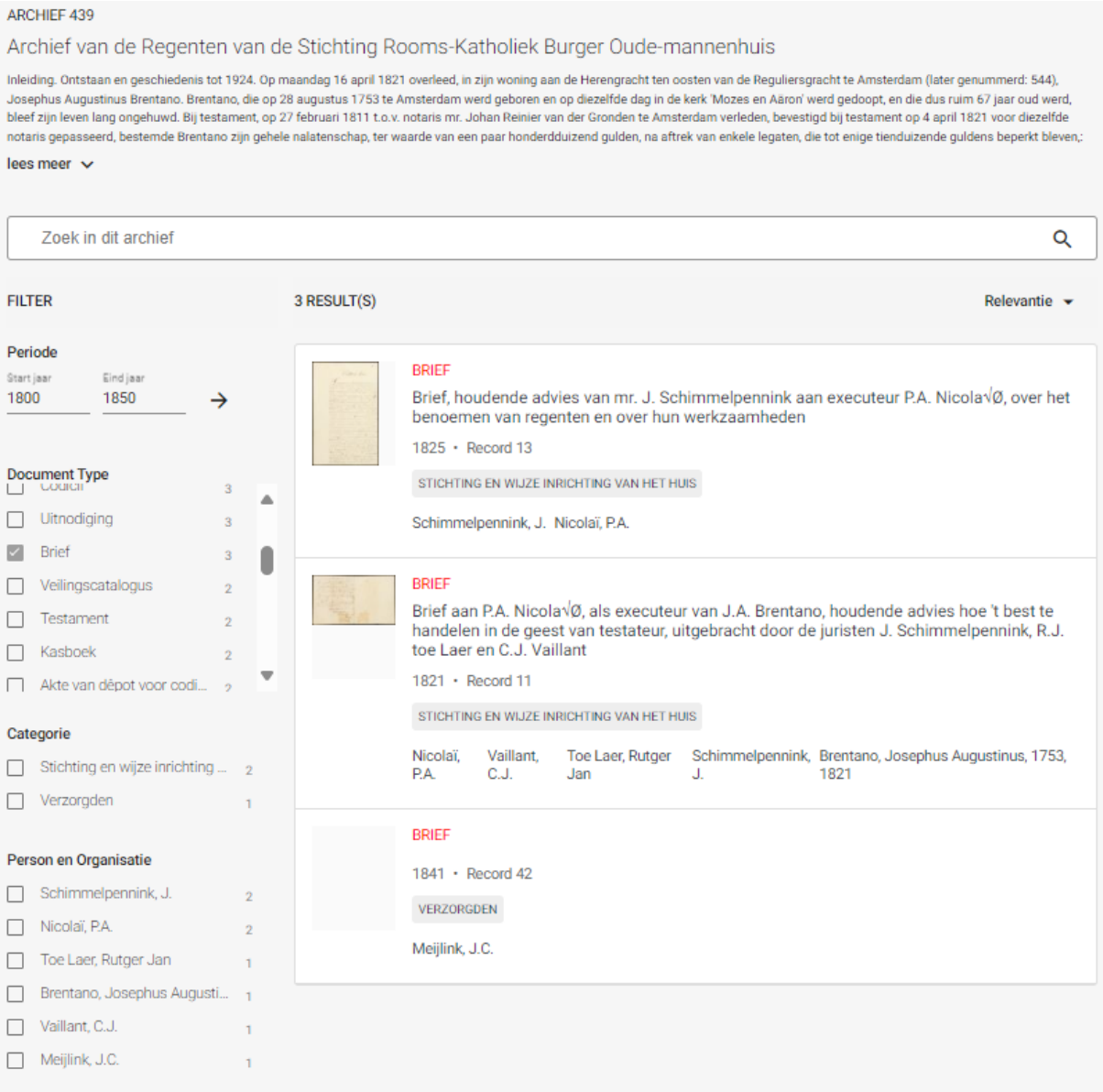
Het tabblad ‘Archiefstukken’ toont een overzicht van alle records binnen het casusarchief. Om de werking daarvan het beste te kunnen beoordelen, is het technisch prototype gemaakt. Dit ziet er als volgt uit:



Rechts zijn de records te zien, of de resultaten van een zoekresultaat daarbinnen. In de zoekbalk daarboven kun je één of meerdere zoektermen plaatsen. Zodra je begint met typen, worden al een aantal suggesties voor zoekopdrachten gegeven, gegenereerd uit de beschikbare gegevens (*autofill*). Je kunt een zoekopdracht ook combineren met filters, links in beeld, enigszins te vergelijken met het zoeken in een webshop. In het prototype staat in een filter achter iedere term de hoeveelheid keren dat deze voorkomt, om overzicht te genereren. De filtertermen passen zich aan als deze gecombineerd worden met andere filters of zoektermen. De termen zijn in het prototype steeds gerangschikt op aantal, de meest voorkomende vooraan. Andere vormen van rangschikking zijn natuurlijk ook mogelijk, bijvoorbeeld op relevantie door een vooraf gedefinieerd ‘gewicht’ mee te geven aan (de plek van) bepaalde termen.

De filtermogelijkheden zijn grotendeels voortgekomen uit de in het vorige hoofdstuk beschreven bewerkingsacties van metadata. Je ziet dat er alleen nog records zijn, geen hiërarchische boom, geen rubrieken. Alle gegevens staan direct bij het record of zijn geplaatst bij contextentiteiten als ‘persoon’. Omdat we *documenttype* en *categorie* als terminologiebron hebben aangemaakt en hebben gekoppeld aan de records, kun je die ook als filter inzetten. Evenals datering en de naam van een persoon of organisatie.

Met deze mogelijkheden kun je makkelijker overkoepelende zoekopdrachten doen, zoals “geef mij alle records waarbij P.A. Nicolaï betrokken is” of “geef me alle brieven tussen 1800 en 1850.” De laatste zoekopdracht is geïllustreerd in onderstaand voorbeeld:



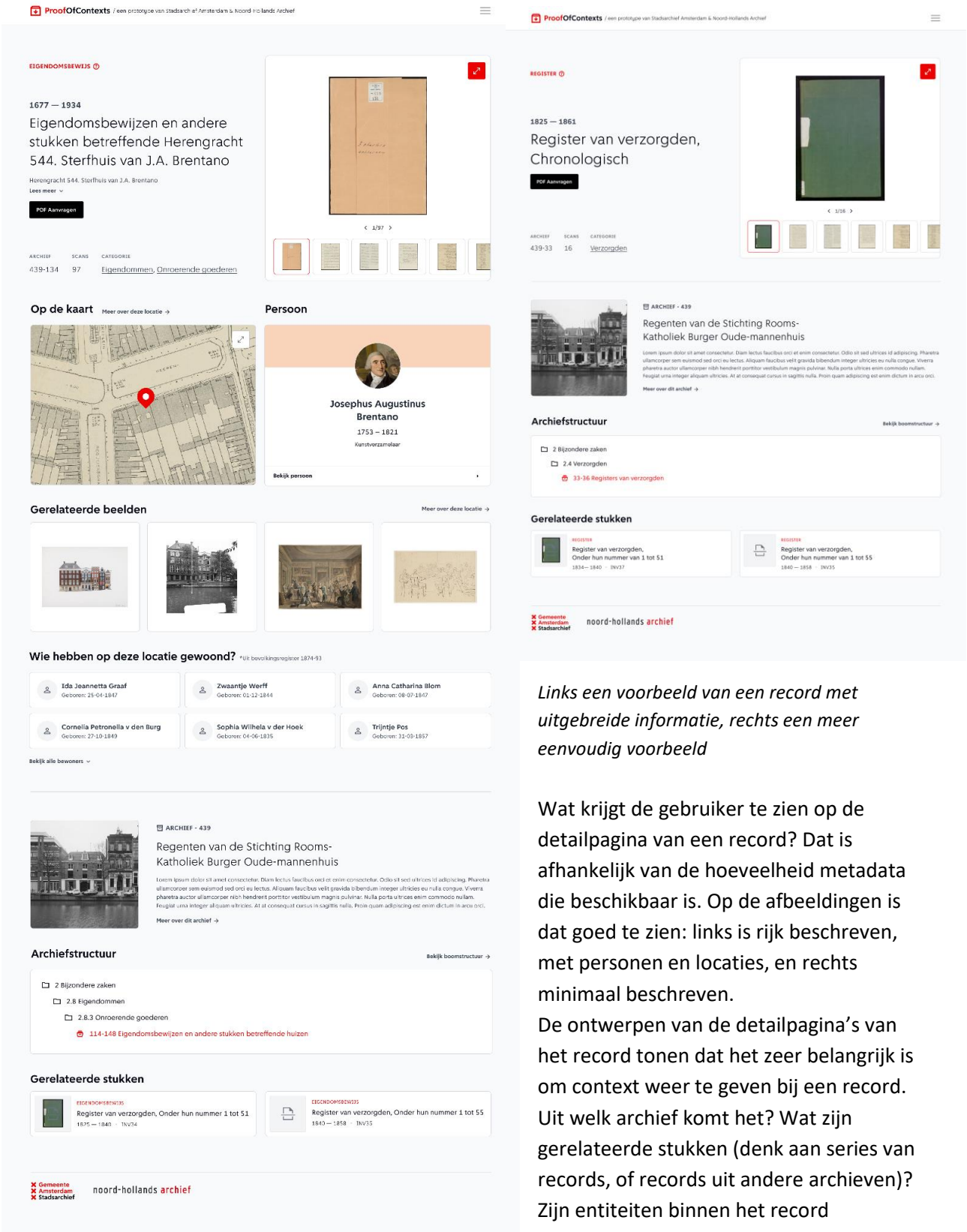
Om deze vraag te beantwoorden, zou de gebruiker in de bestaande inventaris alle rubrieken door moeten gaan om brieven te identificeren. In bestaande websites van archiefinstellingen is het vaak wel mogelijk om te zoeken met zoekwoorden en een datering, maar het resultaat krijg je dan gepresenteerd *in* de boomstructuur. De records kunnen ver uit elkaar staan. Je moet dan best veel

scrollen en klikken. Met 266 records is dat al een klus, laat staan als je het over een archief hebt met duizenden beschrijvingen. In ons prototype krijg je alle records in één oogopslag te zien en kun je deze makkelijker met elkaar vergelijken.

Nu hebben we dit voor één casusarchief gedaan. Maar stel je voor dat het voor alle archieven zo werkt. Dan kun je eenvoudig een zoekopdracht maken als:

- Geef me alle records in de categorie financiën van archief X, Y en Z
- Geef me alle records waarin het woord diamant in de beschrijving staat, en als straat de Jodenbreestraat heeft
- Geef me alle records van archieven van de gemeentepolitie in Nederland met als documenttype ‘dag- en nachtrapport’, tussen 1940-1945

Op deze manier kan een gebruiker meer data-georiënteerd zoeken, op basis van de kenmerken van de records, en niet uitsluitend op basis van zoekwoorden. De gebruiker hoeft niet meer inventaris voor inventaris op zoek te gaan naar records, maar kan zoekresultaten uit alle inventarissen van een archiefinstelling (of zelfs van meerdere instellingen) op één hoop zien. De herkomst van een record (namelijk uit welk archief het afkomstig is) wordt daarmee een contextueel gegeven, en is niet meer uitsluitend het startpunt van de zoektocht.



Links een voorbeeld van een record met uitgebreide informatie, rechts een meer eenvoudig voorbeeld

Wat krijgt de gebruiker te zien op de detailpagina van een record? Dat is afhankelijk van de hoeveelheid metadata die beschikbaar is. Op de afbeeldingen is dat goed te zien: links is rijk beschreven, met personen en locaties, en rechts minimaal beschreven.
De ontwerpen van de detailpagina's van het record tonen dat het zeer belangrijk is om context weer te geven bij een record. Uit welk archief komt het? Wat zijn gerelateerde stukken (denk aan series van records, of records uit andere archieven)? Zijn entiteiten binnen het record geïndiceerd? De gebruiker die vanuit een zoekopdracht op deze pagina komt, kan zo het stuk meteen goed interpreteren, en hoeft niet per se een hele boomstructuur door te ploegen. De gebruiker kan wel nog het 'kruimelpad' zien van de oorspronkelijke inventaris (zie de *Archiefstructuur* in het ontwerp).

Hoofdstuk 4. De reacties

Wat vindt de gebruiker van de ideeën en de ontwerpen van de prototypes? Op drie momenten zijn eerste reacties gepeild: tijdens een tussenpresentatie van het ontwerp aan Noord-Hollandse archivarissen die aan de conceptuele sessies hebben deelgenomen, bij een tussenpresentatie aan de bestuursleden van de Kring van Noord-Hollandse Archivarissen, en in de vorm van een gebruiksonderzoek onder eindgebruikers van archieven.

Tussenpresentatie Noord-Hollandse archivarissen

Na de ontwerpweek met Norday en Spinque zijn de ontwerpen gepresenteerd aan de deelnemers aan de conceptuele sessies, de groep Noord-Hollandse archivarissen. Zij zijn enthousiast over de ontwerpen. Er zijn zeker nog vragen, maar die gaan over vervolgstappen, de realisatie daarvan, en wat je verder nog zou kunnen doen om de toegankelijkheid te vergroten. Het ontwerp *an sich* lijkt niet ter discussie te staan. Enkele reacties van deelnemers aan de conceptuele sessies:
- “Het is pionierend, inspirerend en tegelijkertijd nuttig voor andere archiefinstellingen” - Jeltsje Stobbe, Regionaal Archief Alkmaar
- “Het resultaat was mooi, en een prima startpunt om mee verder te werken lijkt me”- Olga Minkema, Archief Gooi en Vechtstreek

Tussenpresentatie Kring van Noord-Hollandse Archivarissen

Behalve aan medewerkers, is de tussenpresentatie ook voorgelegd aan het bestuur van de Kring van Noord-Hollandse Archivarissen. Net als hun medewerkers, zijn de bestuursleden enthousiast over het ontwerp. Ook hier is vooral verder gepraat over de vervolgstappen en verder onderzoek. De bestuursleden benadrukken dat het een software leverancier-onafhankelijke oplossing is.

Gebruiksonderzoek

Het is natuurlijk mooi dat wij als organisatoren en de deelnemers van de werksessies enthousiast zijn, maar de *proof* is uiteraard *in the pudding*. Daarom zijn vijf mensen uitgenodigd om deel te nemen aan een gebruiksonderzoek, uitgevoerd door Norday. Deze mensen zijn ervaren en beginnende onderzoekers, professioneel en amateur, die eerder actief zijn geweest op archiefwebsites, zodat de oude en nieuwe manier van zoeken goed met elkaar vergeleken kan worden.

Tijdens het gebruiksonderzoek zijn de ontwerpen voorgeschoteld aan de deelnemers. Er is hen gevraagd om hardop te beschrijven wat er te zien is, wat ze verwachtten dat er zou gebeuren als ze ergens op klikten, en om een zoekopdracht uit te voeren in zowel de bestaande zoekinterface als in het nieuwe prototype.

De voornaamste bevindingen zijn, volgens Norday:

De nieuwe manier van zoeken biedt snel resultaat
Ondanks dat mensen gewend zijn aan het doorbladeren van archieven door middel van een boomstructuur, hebben de deelnemers het nieuwe zoeken en filteren snel onder de knie. Het leidt hen eenvoudig naar het juiste resultaat.

- "Het is eigenlijk heel logisch, aangezien je aan de zijkant kan zien naar wat voor soort document je op zoek bent"

De oude manier van zoeken werkt niet intuïtief en erg traag
Met de nieuwe manier van zoeken lukt het iedereen om het juiste document te vinden. Dat is anders in de oude, bestaande zoekinterface, waar het twee van de vijf respondenten niet lukt om precies hetzelfde document te traceren. Mensen die het wel vinden, doen er beduidend langer over dan in de nieuwe opzet. De deelnemers klikken verkeerde mappen open, proberen zoektermen die niet werken, of gebruiken zelfs hun browserzoekfunctie.

- "Dus eigenlijk moet je alles openklappen en overal doorheen gaan om te kijken waar het nou precies staat. Dat vreet tijd, maar dat ben ik gewend, want de meeste archieven werken eigenlijk zo"
- "Ja, ik vond het [in de huidige website, met hiërarchische boomstructuur] een stuk moeilijker. Ik moest nota bene *Ctrl-F* gebruiken om bij eigendomsbewijzen uit te komen, terwijl ik in de vorige omgeving [de nieuwe prototypes] gewoon heel makkelijk kon filteren op 'eigendomsbewijs'"

De oude boomstructuur kan soms gemist worden
De meeste respondenten missen de oude hiërarchische boomstructuur niet in de nieuwe opzet, al geven sommigen aan dat de mappen een gevoel van volledigheid en context kunnen geven. Eén respondent vraagt zich bij de nieuwe opzet af of ze nu alles uit een map heeft gezien. Het gevoel van context zou kunnen worden ondervangen met goede tagging en het duidelijk tonen van gerelateerde records op de detailpagina (zoals 'Andere eigendomsbewijzen').

Het dashboard maakt indruk en biedt overzicht
Het dashboard bij het tabblad 'overzicht' in het visueel ontwerp wordt visueel aantrekkelijk gevonden en kan een handig overzicht bieden als je niet precies weet wat je zoekt.

De verrijking van records maakt enthousiast
Alle testpersonen worden enthousiast van alle doorklikmogelijkheden en gerelateerde content op de detailpagina van een record. Ze zijn hierdoor positief verrast en het maakt de onderzoekers nieuwsgierig naar meer. Met name de mogelijkheid om een kaartweergave te gebruiken en andere bewoners van een locatie te zien, spreekt erg tot de verbeelding.

- "Ik vind dit wel gaaf hoor, want je ziet gewoon die kadastrale kaart waar het huis precies was. (…) En je ziet meer foto's die ermee te maken hebben. En iedereen die er gewoond heeft"
- "Er zijn vast heel veel onderzoeken aan vast te knopen"

Concluderend stelt Norday: “er is veel enthousiasme over het nieuwe zoeken en filteren. Mensen geven aan het directe van de nieuwe opzet te waarderen en ze vinden sneller wat ze zoeken. De oude boomstructuur lijkt te kunnen worden losgelaten, mits het gevoel van context en volledigheid behouden blijft. Dit kan worden opgevangen met duidelijk van elkaar te onderscheiden filters op de zoekpagina en specifiek gerelateerde records op detailpagina’s. Ook het visuele overzicht en de verrijking van records met kaarten en personen valt erg in de smaak en kan onderzoekers verder helpen in hun onderzoek.”

Conclusie en vervolg

De opzet van de Proof of Contexts is geslaagd: we hebben met een casus aangetoond dat de klassieke inventaris daadwerkelijk omgevormd kan worden naar een netwerk van gegevens. Het ontwerp wordt bovendien positief ontvangen door zowel archivarissen als eindgebruikers: het lijkt een meer intuïtieve en gerichte zoekervaring te bieden. Het ontwerp en de metadatabewerkingen die daaraan ten grondslag liggen, zijn een voorbeeld van hoe het zou kunnen worden uitgevoerd. Het nodigt uit tot verdere ideevorming, onderzoek en concretisering in vervolgstappen. Met als stip op de horizon om niet alleen één casusarchief, maar al onze archiefinformatie op een meer intuïtieve en gerichte manier aan te bieden aan de gebruiker.

Vervolgstappen

Er kunnen concrete vervolgstappen genomen worden om de ideeën verder te brengen. Niet alles hoeft in één keer, dat zou ook niet verstandig zijn omdat het te veel is. Het helpt om een stappenplan te maken om te prioriteren, te experimenteren, te leren en bij te sturen. Dit helpt ook om inzicht te krijgen in de benodigde investering in middelen en mensen. Drie concrete vervolgstappen liggen voor de hand:

1. Een pilot om de beoogde techniek voor automatisering te testen

De metadatabewerkingen zijn nu met de hand gedaan, maar kan dit ook deels geautomatiseerd? In het tweede hoofdstuk is een mogelijke methode gepresenteerd, waarin *keywordextractie* en het verrijken van termen centraal staat. Werkt deze methode ook in de praktijk?

Daarbij zijn er wel nog een aantal zaken om uit te zoeken:

- Er zijn ook records die geen categorie hebben, omdat ze in een archief zitten zonder rubrieken. Dit is het geval bij kleine archieven, dan wordt er geen rubriekenstructuur aangemaakt. Hoe erg is dat in de zoekomgeving? Kunnen die records op basis van de gegevens in de beschrijving alsnog aan een categorie worden toegevoegd?
- Uit de lijst van 70.000 rubrieken van het NHA weten we dat er ontzettend veel mogelijke categorieën zijn. Die kunnen *as-is* worden overgenomen, en dan later gekoppeld worden met meer generieke categorieën. De twee labels bestaan dan naast elkaar: de generieke categorie voor gebruik als filter, de specifieke categorie als tag, zodat er geen informatie verloren gaat. Hoe dit praktisch werkt, moet worden uitgezocht.

2. Het ontwikkelen van terminologiebronnen

Om te kunnen verrijken, zijn terminologiebronnen nodig. Deze ontbreken nu voor documenttype, categorie/classificatie, rollen van actoren en organisatievormen. Het beste zou zijn als de archiefsector deze terminologiebronnen gezamenlijk ontwikkelt en beheert, zodat er echt sprake is van interoperabiliteit en gebruikers over archieven en collecties heen kunnen zoeken.

3. Een pilot met meerdere archieven

De prototypes zijn nu gemaakt voor één archief, het Oude-Mannenhuis in Amsterdam, bewust een relatief klein archief in omvang. Maar werkt deze manier van zoeken ook in meerdere archieven en

blijven de conclusies overeind? Hoe ziet het zoekscherm er dan uit, zoekend door meerdere archieven tegelijkertijd? En hoe creëer je daar overzicht?

Er kan bijvoorbeeld een pilot worden gedaan met de gemeentepolitiearchieven van verschillende archiefinstellingen in Noord-Holland. Deze archieven zijn niet alleen qua omvang en structuur heel vergelijkbaar, maar ze worden ook veel gebruikt. Kortom, er kan een grote waardewinst voor de gebruiker worden verwacht.

Aandachtspunten

Naast deze concrete vervolgstappen zijn er nog enkele aandachtspunten en aanbevelingen.

Datamodellen aanpassen
Voordat een instelling tot actie over kan gaan, zal zij eerst haar datamodel moeten aanpassen, want er komen entiteiten en kenmerken bij, en er worden gegevens verhuisd, gekopieerd en toegevoegd.

Datagedreven dashboards
Om overzicht te creëren, zijn in het prototype *dashboards* gemaakt, zo veel mogelijk geautomatiseerd en gedreven door data. Zoals beschreven in hoofdstuk drie is er een keerzijde, want het dashboard kan alleen worden opgebouwd uit nu beschikbare data. Veel data ontbreken nog. Als je bijvoorbeeld wilt laten zien welke records het meest geraadpleegd worden, moeten deze gegevens wel beschikbaar zijn.

Perspectieven verruimen
Er zitten ‘stiltes’ in archieven. Records ontbreken en beschrijvingen zijn gemaakt vanuit het perspectief van de archiefvormer. Het toevoegen van andere perspectieven biedt andere toegangen tot records.

Verzamelbeschrijvingen en indexen
Sommige records hebben een directe relatie met andere records, bijvoorbeeld door verzamelbeschrijvingen of indexen. In het prototype is de navigatie tussen deze records niet uitgewerkt, maar het is wel heel fijn voor de gebruiker, zo bleek ook uit het gebruiksonderzoek.

Jargon uitleggen
Op veel plekken op archiefwebsites is sprake van jargon. Naast dat dit veelal vermeden kan worden, kan al bestaand jargon worden verduidelijkt door termen ter plekke uit te leggen. In het prototype is dit op enkele plekken gedaan door een ‘i’-icoontje toe te voegen, en de betekenis weer te geven als *hoover-over*. Terminologiebronnen helpen hierbij.

Nieuw binnenkomend archief
De PoC handelt vooral over metadatabewerkingen op de zogenaamde *legacy* metadata van archiefinstellingen, archieven die er al decennialang liggen. Maar er wordt voortdurend nieuwe archiefinformatie aangemaakt en verkregen. Er moet een werkinstructie gemaakt worden voor de invoer van nieuw archief dat wordt geïnventariseerd, zodat het meteen kan worden gedaan naar de nieuwste inzichten.

Metadata genereren en verrijken

De PoC laat zien dat de kwaliteit van de metadata enorm belangrijk is, maar de metadata die we graag willen zien is lang niet altijd aanwezig. Kijk maar naar het ontwerp van het record met minimale gegevens. Wat je niet aan metadata hebt, kun je ook niet laten zien, toevoegen of standaardiseren. De creatie van doorzoekbare data kan aanzienlijk worden versneld door analoge records te scannen, transcriberen door middel van HTR/OCR, en de entiteiten daaruit te extraheren, om deze vervolgens te standaardiseren en koppelen. Om de digitale toegankelijkheid van archief- of erfgoedinformatie te kunnen optimaliseren, is een integrale aanpak nodig.

Oplossing bieden voor geen zoekresultaten

Wat krijgt een gebruiker te zien als een zoekterm geen resultaten oplevert? Hoewel deze vraag *out of scope* is geweest in de uitvoering van de PoC, is deze door veel betrokkenen gesteld. De zoekervaring kan verbeterd worden als een gebruiker op basis van de zoekterm suggesties krijgt, bijvoorbeeld om bepaalde archieven te bekijken. Dat kan gedaan worden door een zoektermenanalyse van de zoekomgeving te maken, daar categorieën uit te halen en die als tag toe te voegen aan het archiefblok.

Zoekvragen in natuurlijke taal mogelijk maken

Er zijn experimenten met zoeken door middel van het gebruiken van natuurlijke taal, denk aan hoe chatbots nu functioneren. Die kunnen ook getraind worden op de inhoud van archiefstukken. Dit kan een extensie zijn op een meer klassieke zoekomgeving.

Met dank aan

Proof of Contexts is mogelijk gemaakt door de inzet van vele partijen en personen. Onze grote dank gaat uit naar:

- Het managementteam van het Noord-Hollands Archief en het Stadsarchief Amsterdam
- De Kring van Noord-Hollandse Archivarissen
- Norday
    - Brenda Leensvaart
    - Jacco Ouwerkerk
    - Jasper van der Kamp
    - Tessa Verbij
    - Mischa Heesakkers
    - Jules Verdijk
- Spinque
    - Michiel Hildebrand
- Deelnemers conceptuele sessies
    - Frederiek ten Broeke (Regionaal Archief Alkmaar)
    - Jeltsje Stobbe (Regionaal Archief Alkmaar)
    - Olga Minkema (Archief Gooi en Vechtstreek)
    - Charl Barel (Archief Gooi en Vechtstreek)
    - Anouk Schoemaker (Westfries Archief)
    - Mitchell Helant Muller (Westfries Archief)
    - Emmie Snijders (Gemeentearchief Zaanstad)
    - Antoinet Nijssen (Noord-Hollands Archief)
    - Harco Gorter (Noord-Hollands Archief)
    - Naomi Verlaan (Noord-Hollands Archief)
    - Milan Groot (Noord-Hollands Archief)
    - Mirjam Schaap (Stadsarchief Amsterdam)
    - Anne-Marie Kwakernaak (Stadsarchief Amsterdam)
    - Eric Heijselaar (Stadsarchief Amsterdam)
    - Niek Verhoeff (Stadsarchief Amsterdam)
    - Ilse Rombout (Stadsarchief Amsterdam)
    - Jaap Hoogteijling (Stadsarchief Amsterdam)
    - Herbert Jonker (Stadsarchief Amsterdam)
- Deelnemers technische sessie
    - Rick Companje (zelfstandige, en Het Utrechts Archief)
    - Mark Lindeman (zelfstandige)
    - Antoinet Nijssen (Noord-Hollands Archief)
    - Niek Verhoeff (Stadsarchief Amsterdam)
    - Ivo Zandhuis (zelfstandige)
- Deelnemers gebruiksonderzoek: Batuhan, Claudia, Machteld, Christine en Marc
- Redactie en tekstcorrectie: Liesbeth Oskamp

Fotoverslag

Conceptuele sessies, 3 en 10 juni 2024

Technische sessie, 2 september 2024

Ontwerpweek en tussenpresentatie, 9-16 september 2024