



**internationaal instituut
voor sociale geschiedenis**

Van archief/bibliotheek naar humanities
research infrastructure

Aanvraag vernieuwingsinvestering KNAW

Situatie in 2016

De afdeling Collecties bestaat op dit moment uit goed opgeleide en ervaren mensen, maar hun ervaring en kennis is logischerwijs grotendeels op de fysieke, papieren collectie toegesneden en wordt toegepast op het overgrote deel van de huidige dagelijkse werkzaamheden. Maar in alle aspecten van het werk zorgt de digitale revolutie voor grote veranderingen; aan het verzamelen, het beheren, het ontsluiten en het ter beschikking stellen worden nieuwe eisen gesteld die nieuwe werkprocessen, systemen en andere inrichting van organisatie met zich mee brengen en vragen om andere specialistische kennis en vaardigheden. De afdeling Collecties moet een omslag maken om de impact van de *digital turn* in alle facetten goed te kunnen opvangen.

De afgelopen paar jaar is een begin gemaakt met de omslag, maar het is ondoenlijk gebleken om met het huidige geringe aantal medewerkers en hun expertise en vaardigheden echt vaart te maken. Na de ingrijpende reorganisatie van 2013, waarin 13 fte uit de collectieafdeling verdween, is er te veel werk om mensen echt vrij te maken en te zorgen voor een intensieve kennis- en expertiseontwikkeling bij de huidige staf die kan leiden tot een goede inrichting van de organisatie en de werkprocessen. Daar komt bij dat de gemiddelde leeftijd op de afdeling hoog ligt – binnen nu en 2025 gaat meer dan de helft (17 medewerkers, 11,8 fte) van de medewerkers met pensioen. De kennis en vaardigheden van de huidige medewerkers zijn, hoe waardevol ook, niet toereikend voor de gewenste ontwikkelingen. Bovendien is het hard nodig juist de bestaande kennis en vaardigheden in te zetten voor het beheer en behoud de fysieke collectie en het verwerken van de achterstand.

Investing

Om ruimte te creëren en daarmee de gewenste omslag te kunnen maken is het nodig dat er de komende periode in medewerkers met nieuwe vaardigheden wordt geïnvesteerd. Een constructie, waarbij de pensioneringen een paar jaar voorafgaand aan de daadwerkelijke pensioen datum worden opgevuld, zou hierbij uitkomst kunnen bieden. Wanneer er niet pas bij de pensionering van een huidige medewerker een nieuwe specialist wordt geworven, maar al eerder, kan er een beweging op gang gebracht worden waarmee er over tien jaar een hernieuwde goed ingerichte en digitaal onderlegde afdeling staat. Deze afdeling zal in 2025 in omvang niet groter zijn dan de huidige afdeling, maar heeft tijdelijk tijd en ruimte gekregen om zich te ontwikkelen en te groeien naar een professionele organisatie met de juiste kennis en expertise. De huidige functies – waarbij medewerkers veelal een scala van organisch gegroeide werkzaamheden uitoefenen – kunnen dan volgens een strategisch personeelsplan worden aangepast en toegesneden op het profiel van collectiespecialisten van de toekomst

Binnen het gehele werkproces van collecties kunnen de nieuw te werven collectiespecialisten ingezet worden om de digital turn op het IISG te realiseren. Daarbij gaat het om de volgende functies:

- **Digital Archivists:** Collectiespecialisten die het proces van verwerven, verwerken en beschikbaar stellen van digitale collecties leiden en sturen. Deze medewerkers hebben veel kennis van het digitale collectie proces en digitale duurzaamheid. Zij kunnen het voortouw nemen bij het inrichten en verfijnen van het digitale collecties werkproces. Deze digital archivists zijn, elk met dan wel domeinspecifieke dan wel vakinhoudelijk kennis met name betrokken bij het verwerven en verwerken van digitaal geboren collecties.
- **Specialist collectiesystemen:** Collectiespecialisten die verantwoordelijk zijn voor de inrichting en tooling van de digitale infrastructuur zodat deze optimaal kan worden gebruikt voor de

geautomatiseerde werkprocessen van de afdeling Collecties. Dit is niet enkel een taak voor ontwikkelaars, juist de uitwisseling en nauwe betrokkenheid van collectiespecialisten met goede kennis van digitale methoden en technieken is hierbij onontbeerlijk. Deze specialisten zijn in staat voor deze tools de requirements te schrijven en passen in samenwerking met ontwikkelaars en onderzoekers nieuwe technologieën toe bij de vindbaarheid en presentatie van de collecties.

- **Specialist Data Dissemination:** Collectiespecialisten die het gebruik van de collecties binnen het onderzoek stimuleren en faciliteren, die selecties uit de collecties kunnen maken ten behoeve van onderzoek, relaties in gebruikersnetwerken hebben. Deze medewerkers hebben de expertise om collecties op de diverse platformen, onderzoeksstructuren en services aanwezig te laten zijn en zien kansen voor het gebruik van de collecties. Deze medewerkers zijn ook in staat om fondsen te werven waarmee digitalisering van de analoge collecties kan worden gefinancierd.
- **Medewerker Digitalisering:** Collectiespecialisten die digitaliseringsopdrachten kunnen voorbereiden, zowel in huis on demand als extern uitbestede massadigitaliseringsopdrachten. Deze medewerkers zijn ook verantwoordelijk voor de opname van het materiaal in de digitale depots, aanpassingen in de metadata en de beschikbaarheid van dit materiaal.

Naast deze constructie zou geïnvesteerd moeten worden in het versneld inlopen van de verwerkingsachterstand van de fysieke collecties. De expertise om deze collecties aan te pakken is nu nog aanwezig bij de huidige medewerkers en het zou goed zijn een aantal tijdelijke medewerkers op projectbasis voor langere tijd in te kunnen zetten op het ontsluiten van deze collecties. Voor de ontsluiting van het materiaal in vreemde/niet Europese talen zou specifieke talenkennis kunnen worden ingehuurd. Onder leiding van onze goed opgeleide collega's – voor deze met pensioen gaan – kan de achterstand dan in hoog tempo worden aangepakt, waarbij we mikken op een kilometer per jaar.

Het versneld verwerken van de achterstand zou ook in mentaal opzicht kunnen bijdragen aan een toekomstgerichte ontwikkeling van de afdeling; nu drukt de verwerkingsachterstand van de fysieke collecties nog zwaar op de jaarplanning en geeft medewerkers het gevoel dat er geen tot weinig ruimte is om zich bezig te houden met andere ontwikkelingen.

De situatie in 2025

Verzamelen

De jaarlijkse instroom van IISG-collecties bestaat voor 90 procent uit digitaal materiaal. Het verzamelen van dit materiaal wordt uitgevoerd door een team van goed opgeleide (en voortdurend bij te scholen) experts met inhoudelijke kennis van de ontwikkelingen en spelers in de internationale sociale en arbeidersbewegingen en arbeidsverhoudingen. Het team is in staat om het diverse digitale materiaal, waaronder toenemende uitingen op social media, te selecteren, te acquireren en gereed te maken voor overdracht naar het IISG.

Met de grote vaste archiefvormers zijn afspraken vastgelegd omtrent het aanleveren van het digitaal geboren aanvullingen. Het proces van selectie, overdracht, verwerking en opslag gebeurt in nauwe samenwerking en is zo optimaal mogelijk geautomatiseerd. Materiaal van deze archiefvormers komen automatisch op de servers van het KNAW Humanities Cluster terecht en worden vervolgens via afgesproken protocollen automatisch verwerkt en beschikbaar gesteld.

De periode dat er met de verwerving van het (eerste) digitale geboren materiaal ook nog papieren materiaal mee kwam ligt achter ons. Papieren collecties maken nog maar zelden deel uit van de nieuwe acquisities, het gaat hierbij om ‘vergeten’ aanvullingen uit het pre-digitale tijdperk of unieke onderzoeksdata die geen digitale oorsprong kennen.

Beheer en toegankelijkheid

De infrastructuur van het IISG is opgegaan in die van het KNAW Humanities Cluster en ingericht volgens het erkende referentie model *Open Archival Information System* (OAIS), dat een kader biedt voor de duurzame lange termijnarchivering van digitale informatie en is voorzien van een Data Seal of Approval voor haar gecertificeerde Trusted Digital Repository (TDR).

Dit betekent dat het HuC is aangewezen als duurzame, veilige en betrouwbare bewaarplaats voor digitale data. Organisatorisch en technisch voldoet het HuC aan de wereldwijd erkende standaarden en ISO normen om met digitale data om te kunnen gaan.

Voor het ontvangen van omvangrijke digitale collecties in een verscheidenheid aan soorten (tekst, afbeeldingen, video, audio), bestandsformaten en dragers (floppydisks, tapes, cd's, dvd's, usb-sticks, harde schijven etc) is een aanzienlijk opslagvolume ingericht. Dit volume is een voorportaal waarop machines zijn aangesloten die de diverse (verouderde) dragers kunnen uitlezen of waarop de collectievormer zijn collectie door middel van het *File Transfer Protocol* (FTP) kan neerzetten. Voor alle collecties geldt dat de authenticiteit van de bestanden dient te worden gewaarborgd. Authenticiteit is van cruciaal belang bij digitale collecties; de onderzoeker verwacht van de digitale bronnen dat ze authentiek en betrouwbaar zijn. Hiertoe is een state-of-the-art, technisch hoogstaand, waterdicht, failure proof proces ingericht waarbij de (de inhoud van) het authentieke bestand gefixeerd wordt en gegevens aangaande de authenticiteit worden opgeslagen.

Ten behoeve van de lange termijn opslag zijn de originele digitale collectiebestanden opgeslagen op schaalbare, uitbreidbare, duurzame en veilige opslagvolumes die grote hoeveelheden digitale data aankunnen. Het duurzaam opslaan van de bestanden met behulp van een Persistent Identifier (PID) gebeurt op drie verschillende locaties. Van elk bestand bestaan drie kopieën. Dit garandeert dat wanneer een bestand corrupt raakt, het altijd kan worden gerepareerd met behulp van de andere twee kopieën. De toegang tot de objecten wordt volledig geautomatiseerd afgehandeld door gewaarborgde authenticatie-, autorisatie- en identificatieprocessen.

De digitale collecties met de gerelateerde metadata worden *in kopie* beschikbaar gemaakt via platformen (*dissemination API's*) die de data in internationale standaarden en formaten machineleesbaar aanbieden.

Aan de digitale collecties wordt ontsluitingsinformatie toegevoegd. Deze ontsluitingsinformatie omvat verschillende typen metadata die door de inzet van intelligente multilanguage (e-humanities) tools geautomatiseerd worden verkregen. Door middel van patroonherkenning-tools voor tekst en beeld, die gebruikmaken van de reeds aanwezige domeinspecifieke authorities en ontologieën, wordt er contextuele metadata verkregen waarin de relatie van de collectie tot de omgeving wordt beschreven. Audiovisuele collecties zijn toegankelijker dan ooit door ontsluiting middels spraak- en taalherkenning en transcriptie tools. Daarnaast zorgt toepassing van algorithmes ervoor dat grote collecties niet verworden tot *Dark Archives*: immens omvangrijke digitale archieven waar geen doorkomen aan is en die vanwege privacy gevoeligheid nog minimaal een generatie lang in zijn geheel gesloten zullen moeten blijven. Deze ontsluitingsinformatie is cruciaal voor de toegankelijkheid en daarmee voor de

vindbaarheid van de collecties. Het uitgangspunt is dat de collecties optimaal geschikt worden gemaakt voor onderzoek.

De handmatige ontsluiting van de digitale collecties beperkt zich tot het controleren van de privacy aspecten en het waar nodig bewerken van het resultaat van de geautomatiseerde ontsluiting zodat deze naast machine readable ook human readable is.

Hiermee is de focus en expertise van de collectiespecialisten die zich bezighouden met de ontsluiting van de collecties verschoven van het handmatige ordenen en ontsluiten naar het voortdurend verbeteren en optimaliseren van de geautomatiseerde ontsluiting. De semantische interoperabiliteit van de data en metadata is een continu aandachtspunt. Inhoudelijke kennis van internationale metadatastandaarden, relevante vocabulaires en ontologieën en erfgoed specialistische expertise is hierbij onontbeerlijk.

De papieren achterstand zoals deze 10 jaar geleden bestond is verwerkt. Het fysieke deel van de binnenkomende collecties wordt vrijwel meteen ontsloten en is gereduceerd tot een behapbare normale werkvoorraad. De medewerkers kunnen met behulp van nieuwe ontsluitingstechnieken, kennis en vaardigheden de papieren instroom prima bijbenen.

Alle onsluitingsinformatie is duurzaam, met toekenning van Persistent Identifiers (PID) opgeslagen en wordt beheerd in daarvoor toegeruste informatiebeheersystemen. De opname van de beschrijvende metadata in de collectieregistratiesystemen gebeurt geautomatiseerd met geautoriseerde protocollen (OAI-PMH). Contextuele metadata over een collectie is opgeslagen in een systeem dat de collecties toegankelijk maakt voor intelligente onderzoeksvragen die niet tekst-gebaseerd zijn, maar gebaseerd zijn op de kennis en structuur van de specifieke humanities-discipline. Door gebruik te maken van gestandaardiseerde beschrijvingsformaten (RDF, OWL) worden de collectiedata waar mogelijk gesemantiseerd. De digitale collecties maken deel uit van het veel grotere web van Linked Data en collecties wereldwijd.

Discovery & Delivery

De collecties zijn maximaal ontdekbaar en vindbaar. Gebruikers gaan nauwelijks nog naar catalogi van individuele bibliotheken en archieven, maar maken gebruik van services als Worldcat, ArchiveGrid, APEX of Picarta. De collecties van het IISG zijn up to date aanwezig op deze en andere relevante platformen. Onze standaarden zijn hierop afgestemd. Uiteraard maken de daartoe geschikte collectieonderdelen ook deel uit van nieuwe en bestaande onderzoekinfrastructuren als CLARIN, DARIAH, CLARIAH, en CLIO-INFRA, of publieksgerichte structuren als Europeana, het Geheugen van Nederland, the Internet Archive en Google Cultural Institutes.

De collecties zijn voor de onderzoeker integraal beschikbaar ongeacht een verschil in fysieke, digitaal geboren of gedigitaliseerde verschijningsvorm. De digitale bestanden worden in de context van de collectie én andere collecties gepresenteerd (als [Open] *Linked Data*). Daarbij bieden we intelligente innovatieve zoektechnologie en browsmogelijkheden aan, maar ook visualisaties waarmee bijvoorbeeld de collecties in tijdlijnen en op landkaarten worden uitgedrukt. Conventionele vormen van zoeken en toegang zijn gezien de omvang en de aanwas van de hoeveelheid digitale data ontoereikend gebleken.

De onderzoeker wordt ondersteund bij het uitvoeren van onderzoek, waarbij specifiek eigen onderzoeksvragen worden afgevuurd op relevante beschikbare data. Er zijn services opgezet waarbij onderzoekers de door hen geselecteerde data kunnen laten plaatsen op een *dedicated* en vaak omvangrijk opslagvolume. In deze *virtual research environment* (VRE), de digitale onderzoeksomgeving, kan elke onderzoeker indien gewenst uit de enorme beschikbare hoeveelheid data selecties maken en deze bij elkaar brengen in een eigen datacollectie ten behoeve van het eigen onderzoek. Voor de delivery van in omvang grote bestanden wordt gebruik gemaakt van *webservice keys* waarmee de bestanden via *file transfer protocol* naar de gewenste machine kan worden overgebracht. Naast het afvuren van specifiek eigen onderzoeksvragen kan de onderzoeker eveneens de collectie verrijken door metadata aan te vullen, collectieonderdelen te transcriberen, annotaties aan te brengen of data up te loaden.

Voor de niet-gedigitaliseerde fysieke collecties geldt dat deze collecties eveneens optimaal geschikt zijn voor onderzoek. De collecties zijn goed ontsloten, goed onderhouden en toegankelijk in de studiezaal van het IISG, waar onderzoekers het materiaal kunnen opvragen en bestuderen. De analoge collecties van de overige HuC-instituten zijn hier ook te raadplegen. De dienstverlening in de studiezaal is van hoog niveau en heeft goede bijdetijdse en gebruiksvriendelijke faciliteiten voor de onderzoeker. Collectiespecialisten zetten hun *archival intelligence* in om de onderzoeker zo goed mogelijk te ondersteunen bij het onderzoek in de collecties. De onderzoeker kan indien gewenst zelf digitaliseren of dit waar mogelijk kan laten uitvoeren middels *digitization on demand*. Dit digitaliseren snijdt aan twee kanten, de onderzoeker krijgt de collectie (deels) digitaal beschikbaar, en het gedigitaliseerde materiaal is vrijwel meteen ter beschikking gesteld voor alle onderzoekers.