Pleidooi voor een softwarearchief

JEFFREY VAN DER HOEVEN EN FRANK HOUTMAN

Toegankelijkheid en hergebruik van informatie wordt alom aangemoedigd. De aantoonbare authenticiteit van deze informatie speelt hierbij een grote rol, óók bij informatie in digitale vorm. Maar hoe realistisch is de notie van authenticiteit en toegankelijkheid als de oorspronkelijke hardware en software niet meer beschikbaar zijn? Jeffrey van der Hoeven en Frank Houtman van de Afdeling Digitale Duurzaamheid van de Koninklijke Bibliotheek pleiten voor een softwarearchief en vragen de overheid hiertoe het initiatief te nemen.

Voor het probleem van in onbruik geraakte hardware is inmiddels een oplossing gevonden: emulatie. Dit is specifieke software die de functie van hardware kan nabootsen en het mogelijk maakt zonder oude hardware toch de originele programma's te gebruiken. Het gaat hier niet alleen over software die jaren geleden erg gangbaar was (WP5.1, MS-DOS, dBase, Lotus 1-2-3, etc.) maar ook over bedrijfs- of onderzoeksspecifieke toepassingen die op maat gemaakt zijn en uniek in de wereld. Wanneer die toepassingen niet goed meer werken op een moderne computer, wordt emulatiesoftware ingezet waarmee de informatie in de oorspronkelijke digitale omgeving kan worden geraadpleegd.

Voorwaarde is wel dat over de oorspronkelijke software beschikt kan worden. En daar doet zich een groot probleem voor. Sommige organisaties zullen wellicht een kartonnen doos met oude software bewaren, maar geen enkele organisatie zal beschikken over alle software die zij ooit heeft gebruikt. Laat staan dat er een organisatie bestaat die alle in Nederland uitgebrachte software zoals besturingssystemen en applicaties structureel en systematisch bewaart.

Zonder software kan geen enkele oorspronkelijke computeromgeving worden gereconstrueerd, en juist die omgeving is zo bepalend voor de beleving van oorspronkelijke digitale documenten. Een CAD-tekening waarvan de kleuren anders worden weergegeven dan oorspronkelijk bedoeld, kan al onbegrijpelijk worden. Een website uit 1996, destijds ontwikkeld voor Netscape of Internet Explorer 4.0, ziet er compleet anders uit in huidige browsers. Computermodellen die tien jaar geleden voor MS-DOS werden ontwikkeld zijn zeer waarschijnlijk niet meer correct uit te voeren door nieuwe besturingssystemen. Met andere woorden: de oorspronkelijke softwareomgeving is naast hardware essentieel voor correcte toegang tot overheidsdocumenten, onderzoeksdata en andersoortige digitale informatie.

HENK VAN RUITENBEEK

Waarom bestaat er eigenlijk nog niet zo'n softwarearchief? Ten eerste zijn er technische obstakels, zoals de complexiteit van een softwareomgeving: niet alleen de toepassing moet bewaard worden, maar ook het besturingssysteem, extra drivers, plug-ins, lettertypes, etc., tezamen met de bijbehorende metadata. Een tweede probleem wordt gevormd door wettelijke regels, omdat voor veel software een gebruikslicentie nodig is of omdat Digital Rights Management (DRM) het onmogelijk maakt de data van de originele drager te kopiëren. En dan is er nog het probleem dat software zo divers van aard is: er bestaat niet één versie van MS Office, maar er zijn honderden varianten in verschillende talen, meeteenheden enzovoorts.

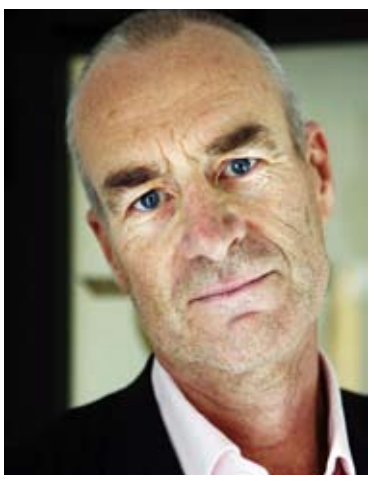
Ondanks dat er geen softwarearchief bestaat, verplicht de Archiefwet iedere overheidsorganisatie de voor de cultuurhistorie relevante (al dan niet digitale) informatie in zijn oorspronkelijke vorm en context te bewaren en over te brengen. Voor informatie op traditionele dragers worden klimatologisch gereguleerde depots gebouwd die beveiligd zijn tegen brand- en waterschade. De toegankelijkheid van digitale informatie is echter niet of nauwelijks gewaarborgd. Om de cultuur en geschiedenis van Nederland ook in digitale vorm voor toekomstige generaties veilig te stellen en toegankelijk te houden is het een absolute voorwaarde dat er een softwarearchief komt. De rijksoverheid zou hierin het initiatief moeten nemen – al is het maar omdat zij als grote partij bij aanbestedingen de mogelijkheid heeft om licenties van softwarepakketten voor dit doel te reserveren.

Alfalab: 'Een virtuele plek om elkaars bestanden en tools te gebruiken'

PETER BOOT

De geesteswetenschappen veranderen niet snel genoeg onder invloed van ICT-ontwikkelingen, onder andere door de kleinschaligheid van veel onderzoek. Een gedeelde digitale infrastructuur kan daarin verbetering brengen, staat in de aanvraag voor Alfalab, een plan dat onlangs door de Koninklijke Nederlandse Akademie van Wetenschappen (KNAW) is goedgekeurd.

Het plan behelst de inrichting van een laboratorium om *ICT-tools* aan digitale bronnen te koppelen, een portaal om de bronnen te ontsluiten, en een 'disseminarium' om de nieuw verworven kennis onder grotere groepen onderzoekers te verspreiden. Het is aangevraagd door een serie KNAW instituten (Huygens, Meertens, DANS, Fryske Akademy, Virtual Knowledge Studio). Ook andere instituten van binnen en buiten de KNAW nemen deel. *e-data&research* vroeg Henk Wals, directeur van het Huygens Instituut en voorzitter van de Alfalab-stuurgroep, om een toelichting.



WIEBE KIESTRA

Henk Wals

Waarom is de voorgestelde investering nodig? Is er iets verkeerd aan de manier waarop de geesteswetenschapper nu werkt?

Wals: Er is niets verkeerd aan de manier waarop de geesteswetenschapper nu werkt, maar informatietechnologie en digitalisering bieden wel nieuwe mogelijkheden. Grote hoeveelheden data en tekst kunnen snel doorzocht worden, conclusies kunnen daardoor worden gebaseerd op meer materiaal, hulpmiddelen voor analyse en visualisatie verschaffen nieuwe inzichten en suggereren nieuwe ideeën. Veel geesteswetenschappers zijn hiermee bezig. Wat zij ontwikkelen levert het nodige op voor hun eigen onderzoek, maar blijft buiten hun discipline vaak onbekend. De geesteswetenschappen zijn nu eenmaal versnipperd. Met Alfalab willen we een virtuele plek scheppen waar onderzoekers uit verschillende disciplines van elkaars bestanden en tools gebruik kunnen maken. Dat is niet alleen efficiënt, maar kan ook tot verrassende nieuwe resultaten leiden.

Wat wordt er mogelijk dankzij Alfalab? Zijn er voorbeelden te geven van hoe de praktijk van het wetenschappelijk onderzoek zou veranderen?

Wals: Simpel gezegd brengt Alfalab bestanden en tools bij elkaar en stemt deze op elkaar af. Onderzoekers kunnen dan kiezen welke ze voor hun onderzoek willen gebruiken. Idealiter kunnen ze dan met een paar muisklikken aan het analyseren slaan. Zelf ben ik historicus. Stel, ik wil een biografie van Willem van Oranje schrijven. Als ik dan in Alfalab zijn gedigitaliseerde brieven aantref en vervolgens een *tooltje* vind waarmee ik chronologisch kan weergeven waar hij verbleef, met wie hij contact had en wat de onderwerpen waren die hij in zijn brieven aanroerde, dan ben ik al een stuk op streek. In één keer heb ik dan het overzicht. Dat scheelt zeeën van tijd. Technisch is dat geen toekomstmuziek meer.

Waarom moet de KNAW hiertoe het initiatief nemen? Is dit niet een nationale aangelegenheid?

Wals: De KNAW staat voor het onderzoek in Nederland en dus ook voor de vernieuwing daarvan. *Computational humanities* gaan de komende jaren een belangrijke rol spelen bij een aantal KNAW-instituten. Het is dus niet zo gek dat uitgerekend van die kant het initiatief komt voor een nieuw stukje onderzoeksinfrastructuur op dit terrein. Maar de eerste fase van Alfalab is niet meer dan een pilot, die wil laten zien hoe het kán werken. De volgende stap is om Alfalab en de financiering daarvan naar een nationale schaal te tillen. Daarvoor zal een breder consortium gevormd moeten worden.

Waarin verschilt Alfalab van infrastructuren als DARIAH of Clarin?

Wals: Tja, het is soms lastig om dat allemaal uit elkaar te houden. DARIAH is een Europese infrastructuur die toegang wil bieden tot geesteswetenschappelijke onderzoeksbestanden. Alfalab wil dat doen voor Nederland, maar legt meer de nadruk op de combinatie van bestanden en tools. Alfalab en DARIAH zullen op elkaar aansluiten. In beide projecten speelt het Nederlandse DANS een voortrekkersrol. Clarin is ook een Europees project dat overlap vertoont met DARIAH en Alfalab, maar is vooral gericht op de linguïstiek. Ook hierin speelt Nederland een belangrijke rol. Er zitten ook Clarin-mensen in de Alfalab-groep en omgekeerd. Het vult elkaar allemaal aan, we concurreren niet.