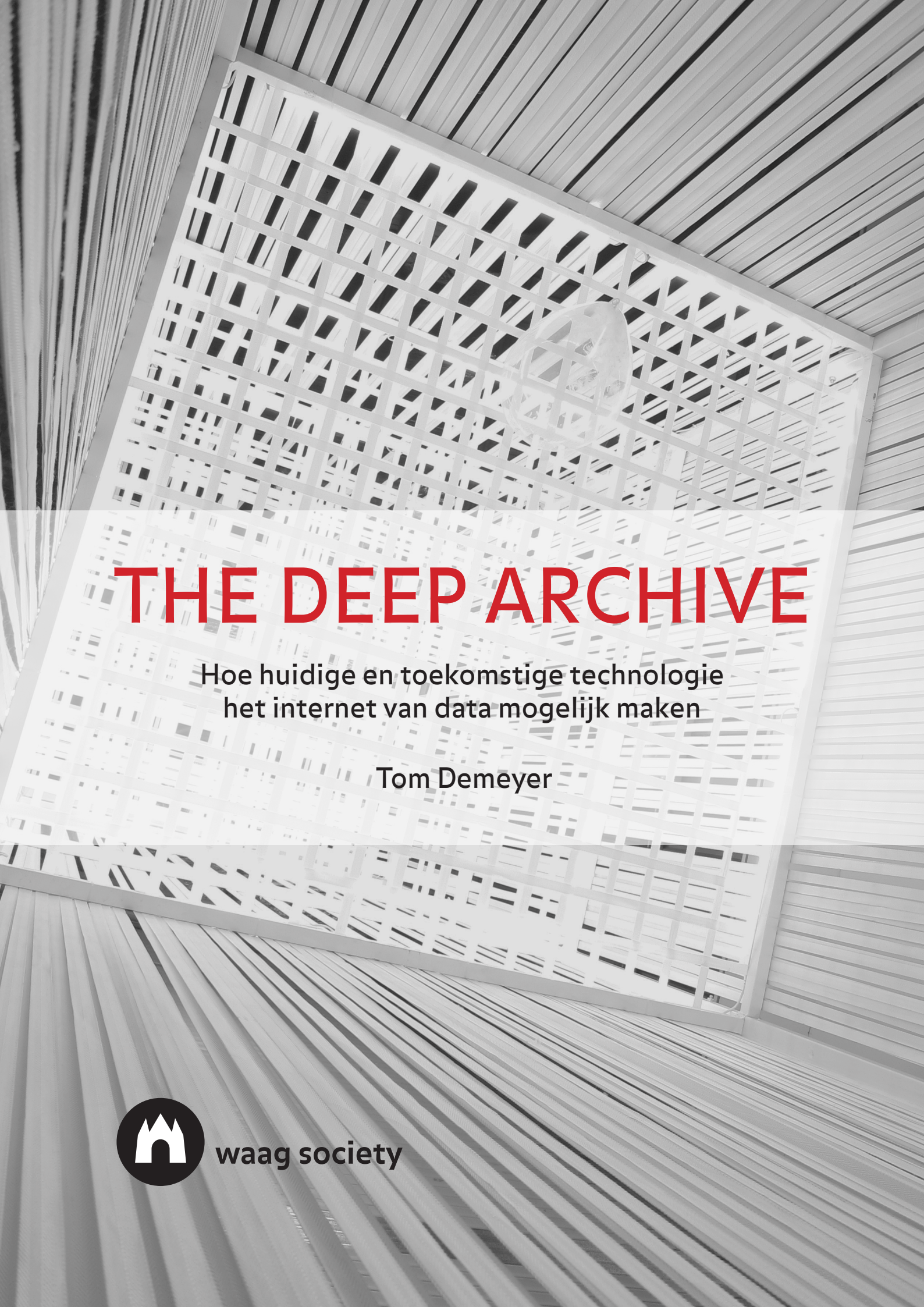


THE DEEP ARCHIVE

Hoe huidige en toekomstige technologie het internet van data mogelijk maken

Tom Demeyer

waag society

The Deep Archive
Hoe huidige en toekomstige technologie het internet van data mogelijk maken
door Tom Demeyer, Head of Technology Development bij Waag Society

Foto omslag: Grid, by Simon Lieschke, 2010 (CC-BY-NC)

Waag Society
Nieuwmarkt 4
1012 CR Amsterdam

waag.org

Inhoud

Inleiding

In 2009 was de totale omvang van alle digitale opslag 500 Exabyte, oftewel 500 miljard Gigabyte, dat is 10 keer een stapel boeken van de aarde naar Pluto. Zeven jaar later is de toename exponentieel en zijn we in het tijdperk van de Zetabyte aanbeland. De stapel boeken rijkt nu tot buiten ons zonnestelsel. Het geeft de magnitude aan van de context van archiefinstellingen. Wat moet worden bewaard en door wie? Het vraagt om nieuwe manieren van werken, van kennis ontwikkelen en kennis delen.

Het vraagt wellicht ook om nieuwe taken voor de traditionele archieven. In samenwerking met Nationaal Archief/Archief2020 en Universiteit Utrecht organiseerde Waag Society in 2016 twee multidisciplinaire *bootcamps*. Hierin identificeerden de deelnemers urgente, actuele vragen en scenario's waarmee de archiefsector zich gaat verhouden tot de uitdagingen van de digitalisering van de samenleving. Hoe draagt de archieffunctie bij aan bedrijfsvoering, democratische verantwoording en cultuurhistorie in het digitale tijdperk? Welke andere archiefspelers hebben het veld betreden en in welke mate hebben zij relevante context en informatie die nodig is om het verleden goed te begrijpen.

Nachtmerriescenario?

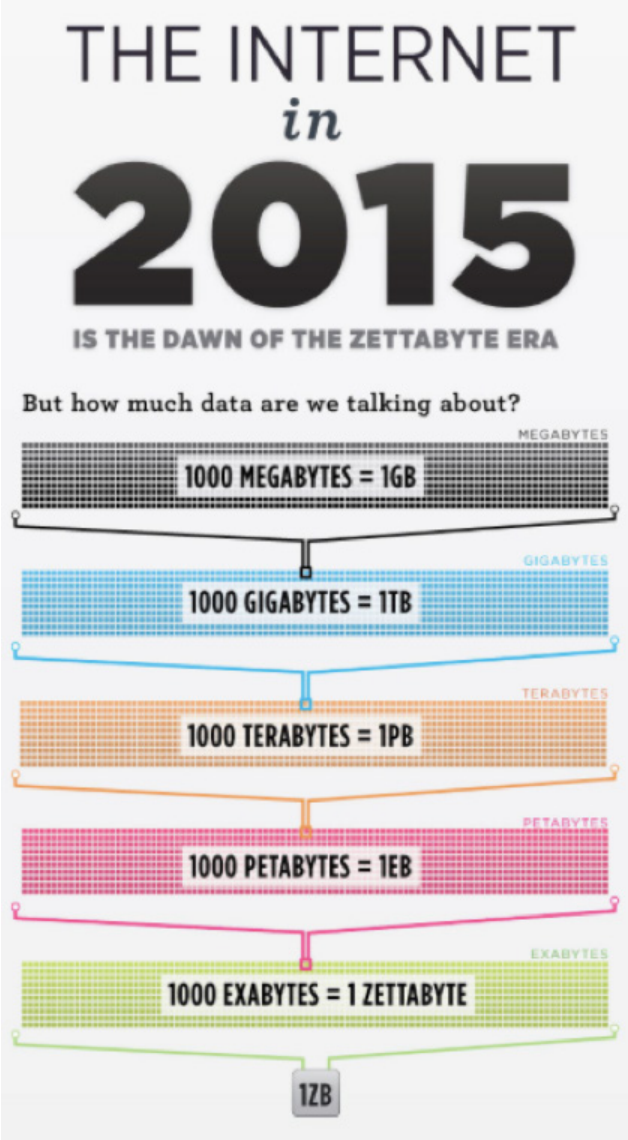
Een van de scenario's die tijdens de bootcamp werd uitgediept heeft betrekking op datamanagement van medische data en dossiers. Dit scenario bevat alle complexiteit die in de centrale vragen ligt besloten. Er zijn dossiers die door de zorgprofessionals worden bijgehouden en mensen houden eigen metingen bij met gezondheid-apps en zelfmeetapparatuur.

Op diverse sociale platforms ligt gedragsdata opgeslagen met gezondheidsinformatie. En er is generieke data die bijhoudt wat medische indicaties, behandelingen en interventies voor gevolgen hebben. Het is in principe mogelijk om alle medische dossiers, eigen metingen met populaire apps en de data van sociale media en koopgedrag te koppelen. Wordt dit een nachtmerriescenario, waarbij mensen gevangen raken in het dataprofiel dat zij ongewild opbouwen of is er een manier om deze data nuttig te gebruiken en de soevereiniteit van de mens in stand te houden?

Uitgestelde selectie

De wereld van archiveren wordt nu beheerst door 'selectie', 'depots' en 'documenten'. Al deze begrippen moeten op de schop: selectie is niet per definitie nodig nu de opslagmogelijkheden onbeperkt lijken te zijn. Een afgebakend depot heeft zijn beperkingen als je voor een goede informatie de data nodig hebt van diverse organisaties, overheidsinstanties, particuliere en private partijen. En daar waar tot dusver volstaan kon worden met documenten zullen de toekomstige gebruikers van 'het' archief behoefte hebben aan de oorspronkelijke data en de algoritmes waar de publicaties op gebaseerd zijn.

Idealiter zouden we de 'selectie' het liefst overlaten aan de toekomstige gebruiker. Alleen zo kan de toekomstig onderzoeker, met haar toekomstige gereedschappen, de huidige maatschappelijke orde optimaal bevragen, of het nu gaat over beleid in de gezondheidszorg, over afnemend vertrouwen in vaccinaties of over grootschalig onderzoek



naar hart- en vaatziekten met behulp van data van persoonlijke fitnesstrackers.

We spreken dus niet alleen over het vraagstuk van het bewaren van data en gegevens, maar ook over de kansen die digitalisering ons biedt om toekomstige generaties te faciliteren in hun onderzoek naar en hun ervaren van het verleden.

De multinationale internetbedrijven en de veiligheidsdiensten hebben een grote technologische voorsprong in het grootschalig opslaan en hergebruiken van data. Zij domineren de discussie rond big data en privacy. De conclusie daarbij lijkt te zijn dat het belang van privacy overschat wordt en dat, als je niets te verbergen hebt, je niets te vrezen hebt van de autoriteiten. De efficiency en veiligheid van het collectief worden hoger ingeschat dan individuele soevereiniteit. De dominantie en zichtbaarheid is zo groot dat het lijkt alsof er geen andere mogelijkheden bestaan en het ons laat twijfelen of we er nog een ander ontwerp mogelijk is van de informatiemaatschappij. Met de centrale vraag of we onze individuele vrijheid inderdaad moeten opgeven om big data mogelijk te maken?

Democratische waarden

Wij geloven dat het van groot belang is om dit speelveld als ontwerper tegemoet te treden met een expliciete opdracht om democratische waarden te verankeren in de technologie. De voortdurende flux van nieuwe technologieën, diensten, wetgeving, curriculumvorming en onderzoek heeft een dergelijk kader nodig heeft. Een gemeenschappelijke visie helpt om de kleinere keuzes waarvoor we dagelijks gesteld worden te maken.

waag.org/archieflab
archief2020.nl

Scenario’s

Op de volgende pagina’s worden twee scenario’s geschetst van verantwoord, open, transparant en inclusief universeel datamanagement, toegankelijk en bruikbaar voor alle mogelijke huidige en toekomstige belanghebbenden. Utopisch, wellicht, maar gestoeld op bestaande technische mogelijkheden of vooruitlopend op niet geheel onrealistische technische ontwikkelingen.

Vanzelfsprekend is technologie niet genoeg; deze ontwikkeling is in eerste instantie een maatschappelijke, en veel van de te bespreken elementen zullen pas ontstaan en ingebed raken wanneer er maatschappelijke noodzaak en draagvlak voor is. Waar aan de ene kant soms ‘grassroots’-ontwikkelingen voldoende zullen zijn, zullen er aan de andere soms ook wettelijke voorwaarden geschapen moeten worden. Het is in ieder geval niet bedoeld als kant-en-klaar systeemontwerp. Specifieke technisch onderdelen zijn niet uitgewerkt en de duivel zit vanzelfsprekend in de vele details.

De scenario’s zijn er uitdrukkelijk op gericht om een denkkader te bieden bij persoonlijke en beleidsbeslissingen op het gebied van datamanagement. Het wil enig gevoel geven voor de bijbehorende issues en complexiteit, en biedt hopelijk ook doorzichten en aanknopingspunten.

Prototype uit de eerste bootcamp

Scenario I. Medisch onderzoek

Het jaar is 2030. Hart- en vaatziekten zijn verdwenen uit de wereld, met uitzondering van een aantal achtergebleven gebieden, zoals bijvoorbeeld Noord-Amerika, waar consensus tussen wetenschap en politiek over aanbevelingen en maatregelen op dit gebied nog op zich laat wachten.

Medische uitdagingen zijn er echter nog voldoende en de huidige 'volksziekte' EWZ houdt laboratoria en universiteiten volop bezig. Liza onderzoekt sociale *drivers* van volksgezondheid en promoveert op een grootschalig onderzoek naar het verloop en de uiteindelijke marginalisering van hart- en vaatziekten in de laatste 50 jaar. Ze heeft hiervoor al enige tijd geleden een onderzoek geïnitieerd bij het 'Deep Archive'. Het protocolleert hierdoor al Liza's interacties betreffende dit onderzoek en kan het zo in een later stadium volledig herhalen, inclusief de relevante data-queries en statistische analyses. Andere onderzoekers kunnen zo Liza's conclusies verifiëren, ook al hebben ze niet Liza's speciale onderzoekstatus op dit gebied.

Liza vindt het wel erg teleurstellend dat ze zo slecht toegang heeft tot gegevens van vóór de tijd van het Deep Archive. Eigenlijk is het vóór 2020 een soort meta-onderzoek, omdat ze alleen toegang heeft tot de geaggregeerde gegevens van onderzoekers uit de betreffende periode. Die gegevens zijn vaak heel beperkt en kleinschalig, bovendien zijn mensen hun activiteiten en *vitals* pas gaan uploaden vanaf halverwege de jaren '10; massaal werd het pas toegepast vanaf 2020 met de komst van het Deep Archive en het maatschappelijk vertrouwen in de zorgvuldigheid van het datamanagement.

Wel heeft ze, vanaf 2020, de beschikking over de gegevens van miljoenen mensen, welke een uiterst compleet beeld schetsen van sport- en fitnessactiviteiten, van ziekte en dieet, rookgedrag en werkomstandigheden. Verbanden met Kamerbesluiten en uitgevoerd beleid, vanzelfsprekend open en geannoteerd beschikbaar in het archief, zijn met behulp van de diensten van Deep Archive snel en navolgbaar gelegd, zo ook die met de discussies op (sociale) media. De conclusies schrijven zich (bijna) zelf.

Scenario II. Delen van data

Het jaar is 2022. Het data-abonnement van Joost verloopt, voor het eerst na de invoering van de algemene uitgebreide data- en archiveringswet. Die wet, relatief haastig ontstaan in reactie op de verrassende opkomst, acceptatie en succes van het Deep Archive, biedt een juridisch kader voor het verzamelen en gebruiken van data tussen overheden, bedrijven, kennisinstellingen en burgers.

Joost, in tegenstelling tot vele anderen, besluit wél zijn persoonlijke databeschikking na te lopen en te kijken waar hij eventueel iets aan wil passen. Het valt mee, eigenlijk is hij het met de meeste instellingen eens. Als gezondheidsfreak besluit hij zijn koopgedrag van etenswaren te koppelen aan zijn medische identiteit. Verder geeft hij aan dat zijn medische gegevens en data van z'n telefoon en activiteitstrackers voor algemeen wetenschappelijk onderzoek gebruikt mogen worden, in plaats van de voorinstelling waar het alleen specifiek medisch onderzoek betreft. De interactieve prompts van het systeem houden alles erg overzichtelijk, gelukkig.

Bij het installeren van een nieuwe hardloop-app op z'n telefoon besluit hij een paar euro te sparen door tijd en afstand van zijn hardloopactiviteiten te delen met de leverancier van de dienst, maar niet de locatie. De data zijn voor de leverancier uitsluitend beschikbaar onder een speciaal hiervoor gecreëerde pseudo-identiteit. Voor wetenschappelijk onderzoek is de hardloopdata beschikbaar onder z'n medische identiteit, dat heeft hij al eerder aangegeven.

In de bovenstaande scenario's moeten bij diverse uitspraken en aannames vraagtekens worden gezet. Om ze technisch, maatschappelijk en bedrijfskundig realistisch te maken is veel ontwikkeling nodig. Onmogelijk is het echter niet. Er is technisch al veel mogelijk en een deel is gebaseerd op een realistische extrapolatie van huidige ontwikkelingen. De grootste horde vormt de politieke sturing. De belangen van verschillende groepen liggen niet op één lijn, en ook ontbreekt het vooralsnog aan een gemeenschappelijke gevoelde urgentie bij grote delen van politiek en maatschappij.

Deep Archive

In de scenario's zitten diverse componenten. De verzameling van deze componenten vormen samen het Deep Archive. De naam Deep Archive is een referentie naar het kunstmatige intelligentie bedrijf Deep Mind dat door Google is overgenomen. Deep Mind onthutste menigeen door in 2016 de wereldkampioen GO te verslaan. De latere kwalitatieve megasprong van de vertaaldienst van Google is een verdere indicatie dat kunstmatige intelligentie nu echte, bruikbare resultaten brengt. Wat dit betekent voor de toekomst is de grote vraag, maar dat het hiermee niet ophoudt lijdt geen enkele twijfel. Het zal zeker een groot effect hebben op het archiefwezen.

De kracht en werkwijze van de huidige kunstmatige intelligentie is gebaseerd op patroonherkenning; op een aantal onderdelen direct van toepassing, maar bij andere is die link niet zomaar te leggen. Aan de andere kant moeten we de kracht van meer regulaire informatietechnologie, zeker in combinatie met kunstmatige intelligentie (KI), natuurlijk niet onderschatten. In deze combinatie gaan de grootste veranderingen plaatsvinden; we zien nu bijvoorbeeld de ontwikkelingen in human computer interactie (HCI) grote sprongen maken. Dit is waar we ons op oriënteren in bovenstaande bespiegelingen: een doorontwikkeling van algoritmische technologieën aan de ene kant, en van KI aan de andere, tot een set van diensten welke samen steeds meer van het Deep Archive-scenario mogelijk gaan maken. In het vervolg zullen we aan de hand van fragmenten uit de beide scenario's reflecteren op verschillende aspecten van het Deep Archive, zowel technisch als maatschappelijk.

... de verrassende opkomst, acceptatie en succes van het ...

We zullen het Deep Archive vermoedelijk nooit als een eenheid zien of ervaren. Anekdotisch is het fijn er een naam aan te geven, maar het zal niet ontstaan doordat een groep wetenschappers, een bedrijf of een overheid op een gegeven moment zoiets op poten zal zetten of aan zal besteden.

Wanneer de voorwaarden daar zijn zullen we het ervaren als een set van praktijken, geboren uit een maatschappelijke behoefte en gedragen en ondersteund door techniek, wetgeving en economische factoren. Dit wil niet zeggen dat we daarin een willoze en machteloze speelbal zijn; uiteraard spelen we allen een rol in het ontwerp van onze maatschappij.

Opslag

Eén van de voor de hand liggende voorwaarden voor het Deep Archive (ofwel het internet van data) is opslag. We willen data opslaan; soms geheim, privé maar deels ook volledig open, met alle gradaties daartussen, en tegelijkertijd deze data te allen tijde toegankelijk houden voor rechthebbenden, veilig voor de toekomst en veilig tegen misbruik.
Dit lijkt misschien een onmogelijke opgave, maar veel van de technologie is beschikbaar als 'redundant, distributed storage' systemen. Bittorrent[1] is een voorbeeld. MaidSafe[2] is een ander welke voet aan de grond probeert te krijgen. De opslagruimte op persoonlijke apparaten en pc's is onderbenut, en wordt ook niet kleiner. Het is niet zo moeilijk te zien dat de plek waar data zich fysiek bevindt steeds minder relevant zal worden, zolang het veilig en bereikbaar is.

Goede annotatie en indexering is de grootste uitdaging; sterke, niet gecompromitteerde encryptie is een andere, mogelijk vooral politieke uitdaging. Principieel is het beeld van één grote robuuste, *redundant storage pool* verdeeld over alle online apparaten geen sciencefiction. Een systeem van *accounting* is ook goed voorstelbaar: hoe meer opslag je ter beschikking stelt (zelf, via apparaten die online zijn, of door een dienstverlener te betalen) hoe meer ruimte je in de pool kunt gebruiken. Dit zou zowel op persoonlijke als op institutionele schaal kunnen werken.

Het succes van het internet is mede te danken aan het gedistribueerde karakter en het gebrek aan centrale macht of governance (hoewel daar natuurlijk een en ander op af te dingen is). De komst van 'de' blockchain laat zien dat er verrassende technische mogelijkheden besloten liggen in het internet, met de potentie echt vernieuwend te zijn. De opslag hoeft dus niet centraal beheerd te worden, of 'van iemand' te zijn; dat kan helpen, met bewezen betrouwbaarheid, bij acceptatie en adoptie.

De manier waarop die opslag benaderd en gebruikt wordt is een ontwikkeling met vermoedelijk veel startups, doodlopende wegen, experimenten en successen. Dat zal het definiëren, maar niet kant-en-klaar opleveren. Beetje voor beetje zouden er elementen kunnen werken en 'blijven hangen', welke dan op een gegeven moment een ecosysteem vormen waaraan we onze data durven toevertrouwen.

Een voorschot op zo'n element nemen we alvast: alle onderdelen van het Deep Archive zijn volledig open toegankelijk voor iedereen. Er is geen sprake van beveiliging van systemen, behalve dat er een limiet is op de hoeveelheid data die kan worden opgeslagen (gekoppeld aan 'echte' rechtspersonen). Alle opgeslagen data is óf volledig open (als in open data) óf alleen versleuteld toegankelijk en zodoende onbruikbaar voor iedereen behalve de rechthebbenden. Autorisatie en toegang tot systemen wordt hierdoor verplaatst naar het ontwerpen van een robuust (persoonlijk) key-managementsysteem. Het ontwerpen en aanbieden van een dergelijk systeem is geen simpele taak, maar een fundamenteel issue dat we in ieder geval aan zullen moeten pakken om te kunnen overleven in een gedigitaliseerde maatschappij; hierover later meer.

Een ander voorschotje op technische ontwikkelingen nemen we door het bepalen van de privacy-gevoeligheid van data uit te stellen tot het moment van uitvoer en gebruik van data, niet bij de opslag (die vindt immers alleen versleuteld plaats). Privacy is vooral in gevaar bij combinatie van data, dat gevaar is soms latent en soms niet onmiddellijk duidelijk. Door bij antwoord op gekwalificeerde vragen de privacyaspecten van de geaggregeerde antwoorden te evalueren win je aan beide kanten. Privacy is beter gewaarborgd en tegelijkertijd zijn onderzoek en innovatie op basis van deze data beter gefaciliteerd. Dit vergt wel enig geloof in de mogelijkheden en kwaliteiten van kunstmatige intelligentie.

… medische identiteit … een speciaal hiervoor gecreëerde pseudo-identiteit …

Wat betekent identiteit in een gedigitaliseerde maatschappij? Die vraag is cruciaal wanneer we het hebben over zorgvuldig datamanagement. Van wie is welke data? Wie heeft er toegang, in welke rol heeft die persoon (bedrijf, dienst, proces, machine) toegang tot die data? Onder welke omstandigheden?

De vragen naar de rol en de omstandigheden duiden er al op dat identiteit wellicht niet

zo'n statisch concept is als we vaak denken. Sterker, identiteit, en ook 'identificeren', zijn begrippen die we in hun huidige vorm in een digitale context eigenlijk niet langer nodig hebben. Het gaat veeleer om de begrippen 'eigenschappen' en 'rechten' (in het Engels: 'attributes' & 'entitlements'). In digitale context bestaat een identiteit uit een bepaalde combinatie van eigenschappen.

Een bepaalde set eigenschappen kan een bepaald recht geven, de eigenschap 'ouder dan 18' geeft bijvoorbeeld het recht om een rijexamen af te leggen of alcohol te bestellen. Verschillende combinaties van eigenschappen leveren dan dus ook verschillende identiteiten op. Er is niet één identiteit, en er is digitaal ook niet een 'echte' identiteit.

Dat wil natuurlijk niet zeggen dat je soms een digitale context niet aan een fysieke persoon (of bedrijf) wilt koppelen. Er zijn vele redenen dit te willen, en ook om hier een rechtssysteem, of zelfs een staat, achter te willen hebben staan. Het hoeft alleen in heel veel van die gevallen niet langer af te hangen van identificatie.

Laten we het voorbeeld nemen van Joost's medische identiteit. Voor Liza is het irrelevant wie Joost is, wat zijn 'echte' identiteit is. Zij wil echter wel in staat zijn verschillende gegevens aan elkaar te kunnen relateren; daarvoor koppelt Joost zijn medische gegevens (en eetgedrag) aan een 'medische identiteit'. Deze bestaat dus uitsluitend uit eigenschappen die Joost hieraan koppelt, en niets meer. Andere eigenschappen van Joost (zijn naam, of Joost's 'zelf', wat dat ook zou zijn), zijn hieruit niet af te leiden.

Rechten

Er is echter maar één fysieke persoon die bij een arts deze gegevens aan zichzelf kan koppelen, i.e. de arts ter plekke kan laten zien dat het lijf dat voor hem zit en die medische identiteit bij dezelfde persoon horen; en dat is Joost. Technisch is dit geen probleem, met de juiste cryptografische mechanismen, uiteraard gebruiksvriendelijk bruikbaar en toegankelijk gemaakt. Tegelijk geeft Joost hiermee toegang tot zijn verzekeringsgegevens (die niet aan z'n medische identiteit zijn gekoppeld).

Kort gesteld, toegang tot gegevens hangt af van rechten of 'entitlements'. Deze verkrijg je, voor bepaalde tijd, door op het juiste moment de juiste serie eigenschappen aan te kunnen tonen. Joost bij de arts als eigenaar van zijn medische identiteit. En Liza's eigenschap bij een universiteit als onderzoeker ingeschreven te staan, een onderzoek te hebben geïnitieerd bij het Deep Archive (waar wellicht nog extra eisen werden gesteld), zijn voldoende om haar in staat te stellen relevante (en dus door de gebruiker beschikbaar gestelde) informatie op te vragen en te ontsleutelen. Liza krijgt dus antwoord op een gekwalificeerde vraag; in het algemeen zal het Deep Archive ruwe brondata alleen aan de feitelijke eigenaar leveren (tenzij het open data betreft). Het uitgangspunt is dat technische ontwikkelingen het mogelijk maken specifieke vragen te stellen, waarop antwoorden in geaggregeerde vorm teruggeleverd worden. Het feit dat het onderzoek volledig wordt geprotocolleerd doet recht aan het principe van herhaalbaarheid van wetenschappelijk onderzoek.

Terzijde: er wordt hier terloops over een paar uiterst complexe zaken rond encryptie heengestapt, zie de korte beschrijving hiervan in het kader op de volgende pagina.

Over encryptie

Wanneer we het hebben over versleutelde data behoeft het geen betoog dat en- en decryptie op de 'client' plaatsvindt en niet op de systemen waar de data wordt opgeslagen. In bovenstaand scenario is dat sowieso al lastig omdat de data niet op één, maar potentieel over meerdere systemen gefragmenteerd zijn opgeslagen. Deze benadering verhoudt zich slecht, echter, tot een aantal van de uitgangspunten zoals die hierboven worden geformuleerd.

- Hoe is een systeem in staat data te aggregeren in response op queries wanneer die data versleuteld is en het 'systeem' zelf geen toegang heeft?
- Hoe kan ik data ontsleutelen zonder dat ik daar specifiek toe gerechtigd ben, maar sinds vandaag door een combinatie van factoren (eigenschappen) wel lid ben van de klasse die daar toestemming voor heeft?
- Hoe blijft data, rechtmatig uit het systeem gehaald, veilig en ontoegankelijk voor derden?
- Hoe wordt data, rechtmatig uit het systeem gehaald, ontoegankelijk als de toestemming tijdelijk was of de aanvrager niet meer gerechtigd is?

Op deze vragen is geen pasklaar antwoord te formuleren, op dit moment. Wel is het zo dat er aan de issues die ze opwerpen actief wordt gewerkt en dat er vorderingen worden gemaakt. Recent is door de Radboud Universiteit Nijmegen onder het begrip PEP[4] (Polymorphic Encryption and Pseudonymisation) een techniek geïntroduceerd welke punt twee adresseert, inderdaad ten behoeve van wetenschappelijk onderzoek, waarbij onderzoekers achteraf inzage kunnen krijgen in versleutelde gegevens, terwijl zij daar op het moment van versleutelen niet in beeld waren als gebruiker.
'Homomorphic encryption'[5] daarentegen, is een techniek die probeert het dilemma van punt één aan te pakken door manipulatie ('berekeningen') toe te kunnen passen op versleutelde data zonder dat deze leesbaar hoeft te zijn. De resultaten liggen vervolgens ook weer in versleutelde vorm voor.

Steganografie, gekoppeld aan een harde (maar eventueel wel pseudonieme) autorisatie kan, met wetgeving, zorgvuldigheid in de hand werken bij gebruikers; ook bij punt drie staan we niet helemaal met de mond vol tanden.

Technieken met betrekking tot gesigneerde en letterlijk wiskundig 'bewezen' code zouden ook nog het gebruik van (lokale) proxysystemen kunnen toestaan welke, zonder opslag van data deze 'on the fly' geschikt maken (aggregeren en versleutelen) voor de gebruiker.

Over mechanismen van autorisatie, implementatie en gebruik van pseudo- en anonimiteit is veel meer te zeggen dan hier mogelijk. Het eID stelsel dat de rijksoverheid van 2013-15 ontwierp bevatte in de oorspronkelijke opzet veel van de mechanismen voorwaardelijk voor een systeem zoals hierboven beschreven. De Radboud Universiteit Nijmegen (RUN) is op het gebied van *attribute-based* autorisatie met 'I Reveal My Attributes' (IRMA) toonaangevend, alsook met de ontwikkeling van benodigde vormen van encryptie (zie bovenstaand kader).

… besluit wél zijn persoonlijke databeschikking na te lopen …

Ook hier wordt veel gesuggereerd dat niet vanzelfsprekend is, en achter deze opmerking gaat een wereld van complexiteit schuil waarvan we hier een indruk zullen proberen te geven. We zien langzamerhand bij de wetgever iets meer serieuze aandacht ontstaan voor nuances en risico's bij de opslag en gebruik van data. Met de onthullingen van Snowden, het einde van 'safe harbour'[3] en door de overbelaste privacy waakhonden lijkt het onderwerp iets makkelijker te agenderen dan voorheen. De technische mogelijkheden worden zo groot dat het steeds vaker ook daadwerkelijk op de agenda terecht zal komen. We permitteren ons vooruit te lopen op een wettelijke regeling rond digitale zelfbeschikking en datagebruik. Alle opslag van te archiveren en persoonsgebonden data, en alle gebruik daarvan lopen voortaan via een Deep Archive.

Databeschikking

Hiervoor krijgt iedereen vanaf de geboorte een persoonlijke 'databeschikking', welke in eerste instantie wordt beheerd door de ouders en, na zekere leeftijd, door de persoon zelf. Hierin wordt onder andere vastgelegd wie bij persoonlijke en verzamelde gegevens kan, en onder welke voorwaarden. Vooruitlopend op een verdere, fundamentelere en meer omvattende digitalisering van de maatschappij, echter, gaat het uiteindelijk om veel meer dan simpel datagebruik; deze beschikking wordt je digitale representatie, datgene waarmee andere digitale zelven in interactie treden, en ook de digitale entiteiten die bedrijven, overheden, en instanties representeren. De digitalisering gaat snel, is onontkoombaar en fundamenteel; onze representatie in het digitale domein zal groter worden dan data alleen; maar die representatie zal wel ons moeten vertegenwoordigen, in al onze persoonlijke en culturele complexiteit.

Bovenstaand vooronderstelt een sterk eID mechanisme zoals eerder beschreven, anders zijn de rechten van het 'wie' en van de digitale representaties niet te benoemen. Het is duidelijk dat niet iedereen evenveel bezig is met data, privacy en digitale representatie. Niet iedereen zal de moeite willen of kunnen nemen zijn of haar databeschikking na te lopen, en specifieke keuzen te maken. De mogelijke gevolgen van verschillende keuzen goed te doorgronden en dan te komen tot een persoonlijke, weloverwogen keuze is erg veel gevraagd.

Instellingen

Dit onderkennende is de databeschikking al ingevuld. Wetgevende, maatschappelijke en culturele krachten komen tot deze standaardinstellingen, welke geenszins statisch zijn, maar ook met de tijd en het veranderende landschap van diensten zullen veranderen. Er zijn drie niveaus waarop invloed wordt uitgeoefend op deze instellingen.

- Wettelijk; hier worden de grenzen vastgesteld waarbinnen instellingen kunnen worden gevarieerd. Een geboortedatum of sofinummer kun je niet veranderen. Bij gericht crimineel onderzoek zijn bepaalde autoriteiten gerechtigd een deel van je gegevens op te vragen. Bij verkoop van alcohol kan de verkoper bewijs van leeftijd opvragen. De status van eigenaarschap van transactiegegevens tussen twee of meer personen wordt hier geregeld. Instellingen en mogelijkheden zullen ook zeker veranderen met de leeftijd van de persoon. Deze wettelijke instellingen worden bepaald door de wetgever, uiteindelijk – of mogelijk internationaal, daar lopen we niet op vooruit.

- Cultureel; er zullen verschillende voorinstellingen zijn welke uiting geven aan een culturele of religieuze identiteit. De ouders zullen een ‘pakket’ kiezen dat aansluit op hun eigen identiteit, in het algemeen. De instellingen blijven natuurlijk binnen de eerder gestelde wettelijke grenzen. Maatschappelijke en culturele stromingen en instituties zullen mogelijk dit soort, wellicht soms hevig bediscussieerde, pakketten ter beschikking stellen.

- Persoonlijk; de enige grenzen zijn de wettelijke. De persoon is vrij zo transparant of ondoorzichtig te zijn als de wet toelaat. Er zullen allerlei krachten inspelen op mensen om bepaalde keuzes te maken en instellingen aan te passen; dit is niet anders dan in het analoge leven, daar zullen we mee om leren gaan.

Hoe het er ook uit zal komen te zien, complex zal het zijn, technisch, maar vooral ook qua functionaliteit en interfacing. Hierbij zal zeker geavanceerde human computer interaction gebruikt worden om vanuit het algemene naar het specifieke te komen.

Gevoel zal moeten worden omgezet in ‘regels’. Een idee als: “ik ben wel op m’n privacy gesteld, maar m’n vriendin mag alles weten” zal bijvoorbeeld op een set regels uitkomen die de persoon in kwestie vervolgens weer worden voorgehouden in hun uitwerking. Zeker als in het vervolg blijkt dat deze ook wel erg op koopjes gesteld is, en graag aanbiedingen krijgt op het gebied van cosmetica zal deze geholpen moeten worden in het afwegen van zijn of haar belangen. Interactief, spraak- en data- en beeldgestuurd... Wie zal het zeggen? Het ontwikkelt zich snel.

Beeld uit de tweede bootcamp

Over Archieflab: Bootcamps

Wat zijn de uitdagingen voor de archieffunctie in de komende vijf jaar? Hoe draagt de archieffunctie bij aan bedrijfsvoering, democratische verantwoording en cultuurhistorie in het digitale tijdperk? In samenwerking met Nationaal Archief/Archief2020 en Universiteit Utrecht organiseerde Waag Society twee multidisciplinaire bootcamps gericht op de toekomst van de archiefsector.

De veranderende context waarin archiefinstellingen opereren, onder andere onder invloed van de digitalisering van de samenleving, vraagt om nieuwe manieren van werken, nieuwe vormen van kennis ontwikkelen en kennis delen, maar leidt ook tot nieuwe taken voor de traditionele archieven.

Het doel van de bootcamps was om urgente, actuele vragen te identificeren en daaruit voortvloeiende innovatietrajecten te adresseren waarmee de archiefsector zich gaat verhouden tot de digitale samenleving van de 21e eeuw. In de kick-off van het project werden meerdere onderwerpen gesignaleerd om uit te werken, zoals automatische metadatering, authenticiteit in combinatie met block chain-technologie, standaardisering, informatieveiligheid, en de betekenis van selecteren en vernietigen in een digitaal tijdperk. De bootcamps waren meerdaagse, multidisciplinaire werksessies waarin deelnemers hands-on samenwerkten op actuele thema's.

In de bootcamps werden toekomstige producten en processen als prototypen verbeeld. Deze ontwerpbenadering helpt in het definiëren van de waarden waarop het Innovatielab gebaseerd is, het doorgronden van onderliggende principes, het verbeelden van toekomstscenario's, het bespreekbaar maken van effecten en het bieden van handelingsperspectief.

Noten

[1] https://en.wikipedia.org/wiki/BitTorrent
[2] https://maidsafe.net/
[3] https://nl.wikipedia.org/wiki/International_Safe_Harbour_Principles
[4] https://eprint.iacr.org/2016/411.pdf
[5] https://en.wikipedia.org/wiki/Homomorphic_encryption

waag society