

Reader E-depot

Versie 1.1,

24 mei 2018.

Inhoudsopgave

Inleiding.....	3
1. Wat is archief?.....	4
1.1 Wat is digitaal archief?	5
1.2 Soorten digitaal archief	5
1.3 Bruikbaarheid van digitaal archief.....	6
1.4 Kwaliteitskenmerken van digitale archiefstukken	6
1.5 Overbrenging en uitplaatsing	7
1.6 Oefeningen	7
2. Wat zijn metadata?	9
2.1 Context, inhoud en structuur	10
2.2 Technische metadata	11
2.3 XML.....	11
2.4 Oefeningen	12
3. Wat is een e-depot?	14
3.1 Functies van een digitaal archiveringssysteem	15
3.2 Kenmerken van digitale archiveringssystemen.....	16
3.3 OAIS	17
3.4 Oefeningen	20
4. Proces van digitale archivering.....	21
4.1 Impactanalyse aansluitingen	21
4.2 Procesbeschrijving overbrenging of uitplaatsing	23
4.3 Oefeningen	29
5. Beheer van digitale archieven	30
5.1 Beheer van metadata van digitale archieven.....	30
5.2 Beheer van de digitale bestanden.....	30
5.3 Verwijderen van opgenomen archieven	30
5.4 Oefeningen	31
6. Beschikbaar stellen van digitale archieven	32
6.1 Beschikbaar stellen via een inventaris	32
6.2 Beschikbaar stellen als open data	33
6.3 Beperkingen aan online beschikbaar stellen.....	33
6.4 Oefeningen	33
7. Nieuwe vormen van informatie	34

Inleiding

Deze reader beschrijft de lesstof van het vak 'E-depot' dat in 2018 gegeven wordt aan de studenten van de opleiding Bachelor in Archiving van de Inter-Continental University of the Caribbean in Willemstad, Curaçao.

De lesstof begint met een korte herhaling van kennis die bij eerdere vakken in deze opleiding is opgedaan: wat is (digitaal) archief, wat zijn (digitale) metadata, wat is een (digitaal) archiverings-systeem en wat zijn de kwaliteitseisen die door de ISO15489-standaard worden aangegeven. Vervolgens wordt de brug gemaakt naar een e-depot, een digitaal archiverings-systeem dat gespecialiseerd is in het duurzaam toegankelijk houden van digitaal archief.

Nadat de basis voor digitaal archiveren is gelegd, wordt het proces van digitale archivering behandeld, waarbij het proces van aansluiten van een informatiesysteem bij een archiefvormer op een archiverings-systeem wordt doorlopen. Hierbij worden de instrumenten 'Impactanalyse aansluitingen' en 'Beslisdocument' ingezet, met behulp van het Nederlandse metadataschema TMLO en het bijhorende machine-leesbare ToPX. De lessen worden afgesloten met een klein kijkje in de toekomst van digitaal archief.

Elk hoofdstuk eindigt met een aantal oefenopgaven. De oefenopgaven vragen vaak aan de studenten om voorbeelden te zoeken in hun eigen werkomgeving om de brug te slaan tussen de theorie in deze reader en de praktijk waarin de studenten tijdens hun werk mee te maken hebben.

Jeroen van Luin,
Den Haag,
Mei 2018.

1. Wat is archief?

Archief is (een verzameling van) procesgebonden informatie. Dat wil zeggen: archief is vastgelegde informatie die gemaakt of ontvangen is als onderdeel van een werkproces om vanuit datzelfde werkproces of vanuit andere werkprocessen te kunnen worden bevraagd¹. Het maakt daarbij niet uit of die informatie als invoer dient voor het werkproces, of als uitvoer opgeleverd wordt door het werkproces, of dat het alleen intern, binnen het werkproces zelf is gebruikt.

Het vormingsproces van archief loopt als volgt²:

1. **Waarnemingen** leiden tot **kennis** bij diegene die de waarneming doet;
2. **Kennis** leidt tot **informatie** door het te delen en erover te communiceren;
3. **Informatie** leidt tot **archiefstukken**, door de informatie vast te leggen;
4. **Archiefstukken** leiden tot **archief**, door ze duurzaam te bewaren.

Archief heeft documentaire waarde, het toont aan dat een proces bestaan heeft, en hoe het gelopen is en heeft bewijskracht. Archiefstukken zijn niet opzichzelfstaand, maar hebben relaties met andere archiefstukken en hebben een context: het wie, wat, waar, wanneer en waarom van het archiefstuk.

Voor de vraag of geschreven informatie 'archief' is, maakt het niet uit wie het werkproces uitvoert. Overheid of particulier, ministerie of student: wanneer informatie is opgeschreven en verbonden is met een werkproces, dan is het archief.

Dat iets archief is, betekent niet automatisch dat het ook voor blijvende bewaring bedoeld is. Afhankelijk van de rol die een informatieobject speelt in het werkproces moet het soms al na seconden worden vernietigd, of moet het dagen, weken of jaren bewaard blijven voordat het vernietigd moet worden. Het deel van een archief dat echt voor altijd bewaard moet worden is meestal maar een klein deel van het totale archief.

De noodzaak om een archief aan te leggen en informatie voor een bepaalde tijd te bewaren komt voort uit verschillende redenen. Vanuit de Belastingwet moeten particulieren en bedrijven een administratie aanleggen om belastingaangifte te kunnen doen, en de stukken hiervan een aantal jaren bewaren. Diezelfde stukken kunnen ook nodig zijn om te weten hoeveel voorraad een bedrijf nog heeft en om te weten wie zijn rekening nog niet betaald heeft. Artsen hebben patiëntendossiers nodig voor de behandeling van hun patiënten, en voor de financiële afhandeling van de behandeling. Het bonnetje dat je krijgt bij de aanschaf van een elektronisch apparaat is tegelijkertijd het aankoopbewijs en het garantiebewijs.

Deze redenen worden samengevat in het volgende lijstje:

1. Gebruik in de eigen dagelijkse bedrijfsvoering;
2. Aflegging van verantwoording;
3. Recht- en bewijsvinding;
4. Cultuurhistorische waarde, het persoonlijke en het gemeenschappelijke geheugen.

Voor de overheid zijn er wetten en regelingen³ waarin de rechten en plichten ten aanzien van archieven wordt geregeld. Hierin staat onder andere dat overheidsorganen verplicht zijn om de

¹ Uit: Peter Horsman, "Archiveren, een inleiding", 2011.

² Uit: Laura Millar, "Archives: Principles and practices", 2010.

archieven die ze aanleggen in goede, geordende en toegankelijke staat te bewaren⁴, van alle categorieën informatie te weten hoe lang het bewaard moet worden⁵ en wie het mag inzien⁶, en dat de informatie die blijvend bewaard moet worden na verloop van tijd naar een archiefbewaarplaats moet worden overgebracht⁷.

1.1 Wat is digitaal archief?

Digitaal archief is archief, maar dan digitaal. Dat klinkt flauw, maar geeft wel de essentie weer: conceptueel gezien is er geen verschil tussen digitaal archief en papieren ('analoog') archief. In de wetgeving is dat ook expliciet vermeld: 'ongeacht de vorm'⁸. Het gaat om procesgeboden informatie die door de archiefvormer is ontvangen of gemaakt en is vastgelegd. Alleen is de informatie nu digitaal vastgelegd in plaats van op papier of perkament. Digitaal archief heeft net als papieren archief inhoud, context, vorm, functie en relaties. Ook van digitaal archief moet worden bepaald hoe lang het bewaard moet worden en wie het mag inzien. En ook digitaal overheidsarchief dat voor blijvende bewaring in aanmerking komt moet op termijn worden overgebracht naar een archiefbewaarplaats.

Natuurlijk zijn er ook grote verschillen tussen digitaal archief en papieren archief. Waar inkt echt vast zit aan papier, zit digitale informatie zit niet vast aan de drager (harde schijf, usb-stick, cd-rom, etc). De informatie die in een digitaal bestand zit opgeslagen heeft ook weer hardware en software nodig om gebruikt te kunnen worden. Een digitaal bestand kan gekopieerd worden en zal dan een exact duplicaat hebben, waarvan later niet gezegd kan worden welke nu 'origineel' is, en welke 'kopie'.

Dat de informatie los staat van de drager betekent ook dat de informatie makkelijk via een digitaal netwerk verspreid kan worden. Door informatie via internet toegankelijk te maken kunnen geïnteresseerden van over de hele wereld bij de informatie, gewoon vanuit hun eigen huis. Dit levert nieuwe kansen en nieuwe risico's op: informatie is heel gemakkelijk te delen, maar eenmaal gedeeld is het ook vrijwel niet meer permanent te verwijderen.

1.2 Soorten digitaal archief

We onderscheiden twee soorten digitaal archief: gedigitaliseerd archief en digitaal gevormd archief, meestal aangeduid als 'born-digital' archief.

Gedigitaliseerd archief

Onder 'gedigitaliseerd archief' verstaan we alle digitale archiefstukken waarbij de informatie oorspronkelijk op papier is gevormd, en waar door een scanner of camera een digitale afbeelding van is gemaakt. Wanneer na digitalisering de papieren originelen zijn vernietigd, dan spreken we van 'vervanging' of 'substitutie' en krijgen de digitale reproducties de status van 'origineel'⁹. Blijven de papieren originelen bewaard, dan krijgen de reproducties de status van 'raadpleeg-kopie' en houdt het papier de status van 'origineel'.

³ Voor Nederland: de Archiefwet 1995.; voor Curaçao: de Archieflandsverordening 2007.

⁴ In zowel de Archiefwet als de Archieflandsverordening in artikel 3.

⁵ Archiefwet: artikel 5; Archieflandsverordening: artikel 7.

⁶ Archiefwet: artikelen 14, 15, 15a en 16; Archieflandsverordening: artikelen 19, 20 en 21.

⁷ Archiefwet: artikelen 12 en 13; Archieflandsverordening: artikelen 17 en 18.

⁸ Archiefwet en Archieflandsverordening: artikel 1 lid c onder 1.

⁹ Archiefwet: artikel 7; Archieflandsverordening: artikel 8.

Born-digital archief

We spreken van ‘born-digital archief’ wanneer de archiefstukken digitaal gevormd zijn, bijvoorbeeld in een tekstverwerkingsprogramma, als spreadsheet, digitale presentatie of een e-mailprogramma. Ook informatie die niet in documenten zit, maar als digitale dataset is opgeslagen, geldt als born-digital. Denk hierbij bijvoorbeeld aan resultaten van meetapparatuur of databases met persoonsgegevens die door een bedrijf of overheidsinstelling worden bijgehouden.

Hybride archief

Alhoewel er ook binnen overheden steeds meer digitaal gewerkt wordt, wordt papieren archief nog steeds gevormd. Vaak komt het voor dat een archief deels uit papieren archiefdocumenten bestaat, en deels uit digitale archiefobjecten. We spreken dan van een ‘hybride archief’. Hierbij kan het voorkomen dat archief deels uit papieren dossiers bestaat en deels uit digitale dossiers, of dat een dossier bestaat uit papieren documenten en deels uit digitale documenten. Het komt zelfs regelmatig voor dat een document digitaal gevormd wordt, en daarna uitgeprint wordt omdat het met pen ondertekend moet worden. De digitale en de papieren versie zijn in principe hetzelfde document, maar in een ander ontwikkelingsstadium. Als er een vervangingsbesluit is, komt het zelfs regelmatig voor dat de ondertekende papieren versie wordt gedigitaliseerd en de papieren versie wordt vernietigd. Er is dan dus een born-digital niet ondertekende versie en daarnaast een gedigitaliseerde ondertekende versie.

1.3 Bruikbaarheid van digitaal archief

Eén van de grote voordelen van digitaal archief boven papieren archief is de betere bruikbaarheid. Hiervoor werd al genoemd dat digitaal archief via digitale netwerken gedeeld kan worden, maar ze kunnen ook makkelijker gevonden worden. Born-digital documenten en gescande documenten die voorzien zijn van OCR¹⁰-data kunnen *full-text* doorzocht worden en zijn zo veel nauwkeuriger ontsloten dan papieren archief. Voorwaarde hierbij is wel dat de gebruiker van het archief weet welke terminologie door de opsteller van het archiefstuk is gebruikt of dat de gebruikte zoekmachine in staat is om ook te zoeken op alternatieve termen. Het *full-text* doorzoeken van teksten is voornamelijk dan ook een aanvulling op de normale archivalistische beschrijvingen, en geen vervanging van.

1.4 Kwaliteitskenmerken van digitale archiefstukken

Om hun functie voor bedrijfsvoering, verantwoording, bewijsvoering of herinnering goed te kunnen vervullen moeten archiefstukken, en dus ook digitale archiefstukken, over de juiste kwaliteitskenmerken beschikken. Deze worden in de NEN-ISO 15489-1:2016 genoemd:

1. Authenticiteit

Een authentiek archiefstuk is een archiefstuk waarvan bewezen kan worden dat het is wat het zegt te zijn, dat het is gemaakt door de actor die beweert het te hebben gemaakt, en dat het is gemaakt of verzonden is op het moment waarop het zegt gemaakt of verzonden te zijn;

¹⁰ Optical Character Recognition, waarbij er automatische tekstherkenning heeft plaatsgevonden. Deze herkende teksten kunnen gebruikt worden bij het doorzoeken of beschikbaar stellen van de scans.

2. **Betrouwbaarheid**

Een betrouwbaar archiefstuk is een archiefstuk waarvan de inhoud een volledige en nauwkeurige weergave van de transacties, activiteiten of feiten is waarvan het getuigt. Een archiefstuk opgesteld door iemand met directe kennis van de feiten is een betrouwbaarder archiefstuk dan een stuk opgesteld door iemand die de kennis alleen van horen zeggen heeft;

3. **Integriteit**

Een integer stuk is een stuk dat alleen geautoriseerde veranderingen heeft ondergaan, en waarvan de veranderingen goed gedocumenteerd zijn;

4. **Bruikbaarheid**

Een bruikbaar archiefstuk is een archiefstuk waarvan de vindplaats bekend is en die door geautoriseerde belanghebbenden binnen redelijke termijn en met redelijke inspanning gevonden, weergegeven en geïnterpreteerd kan worden.

1.5 Overbrenging en uitplaatsing

In de archiefwetgeving staat dat alle te bewaren archiefstukken uiterlijk na 20 jaar moet worden overgebracht naar een archiefinstelling. Na overbrenging gaat het zorgdragerschap van het archief over op de minister die verantwoordelijk is voor de archiefinstelling¹¹.

Voor digitale bestanden is 20 jaar een erg lange tijd. Digitale ontwikkelingen gaan zo snel dat veel eerder dan die 20 jaar al beheerhandelingen nodig zijn om de toegankelijkheid van de informatie te kunnen garanderen. En voor de digitale archiefstukken die op termijn vernietigd moeten worden maar die wel lang bewaard moeten worden, zijn diezelfde beheerhandelingen nodig om de stukken duurzaam toegankelijk te houden. Wanneer archiefvormers niet zelf voorzieningen willen inrichten om te zorgen dat hun digitale archief goed blijft, dan kunnen zij kiezen om hun archieven uit te plaatsen.

Bij uitplaatsing blijft de archiefvormer verantwoordelijk voor de zorg over de digitale archiefstukken, beslist zelf over wie inzage mag hebben, en beslist wat er bewaard blijft en wat er op termijn vernietigd wordt. De daadwerkelijke uitvoering van het beheer over de archiefstukken wordt uitgevoerd door iemand anders. Dat kan een commerciële dienstverlener zijn, of een archiefinstelling. Uiteindelijk komt de termijn van 20 jaar in zicht, en moet het te bewaren deel van het uitgeplaatste archief alsnog worden overgebracht van de uitplaatslocatie naar de archiefdienst.

1.6 Oefeningen

Opgave 1.1:

Bedenk 2 voorbeelden van documenten die je in jouw eigen werk ontvangt.

Komen deze documenten vooral op papier, digitaal of allebei binnen?

Zijn ze voor blijvende bewaring, of moeten ze na een bepaalde tijd vernietigd worden?

Worden binnengekomen papieren documenten gedigitaliseerd? Zo ja, wat gebeurt er met de papieren originelen? Zo nee, zou digitalisering van de documenten jouw werk makkelijker maken of niet?

Met welk doel worden deze documenten bewaard?

¹¹ Voor Nederland: de Minister van Onderwijs, Cultuur en Wetenschap. Voor Curaçao: de Minister van Constitutionele en Binnenlandse Zaken

Opgave 1.2:

Bedenk 2 voorbeelden van documenten die je in jouw eigen werk maakt met als doel deze aan iemand anders te sturen.

- a) Bewaar je een kopie van het uitgezonden document, en zo ja, waar worden ze bewaard, en met welk doel worden ze bewaard?
- b) Wordt het document digitaal verstuurd, of wordt het geprint en op papier verstuurd?

Opgave 1.3:

Bedenk 2 voorbeelden van documenten die alleen intern in jouw eigen werkproces worden gebruikt.

- a) Zijn dit ook archiefstukken? Zo ja, waarom, zo nee, waarom niet?
- b) Waar zouden dergelijke documenten moeten worden bewaard?
- c) Met welk doel zouden deze documenten moeten worden bewaard?

Opgave 1.4:

Bedenk zijn 2 voorbeelden van documenten die je tijdens jouw werk tegenkomt, maar die géén archiefstuk zijn. Geef daarbij aan waarom het volgens jou geen archiefstukken zijn.

Opgave 1.5

Archiefstukken kennen de kwaliteitskenmerken authenticiteit, betrouwbaarheid, integriteit en bruikbaarheid.

- a) Geef een voorbeeld van een document uit jouw eigen werkomgeving waarbij je kunt aantonen dat de authenticiteit, betrouwbaarheid, integriteit en bruikbaarheid kloppen.
- b) Verzin een voorbeeld van een document dat wel authentiek, integer en bruikbaar is, maar toch niet betrouwbaar.
- c) Verzin een voorbeeld van een document dat wel betrouwbaar, integer en bruikbaar is, maar niet authentiek
- d) Horen documenten die niet authentiek of betrouwbaar zijn in een archief?

2. Wat zijn metadata?

Metadata zijn data over data, of gegevens over gegevens. Volgens de NEN-ISO 15489-1:2016 norm:

gegevens die context, inhoud en structuur van archiefbescheiden en hun beheer door de tijd heen beschrijven

De metadata moet iets zeggen over de bedrijfscontext van een archiefstuk of set van archiefstukken, en over hun afhankelijkheden onderling en met andere archiefstukken, over hun juridische en maatschappelijke contexten en over de relaties met degenen die de stukken gemaakt hebben, ze beheren en die ze gebruiken.

Bij papieren archief zijn de metadata handmatig aangebracht. Via nietjes, paperclips, stempels, post-it's, etiketten en/of geschreven of getypte aanwijzingen wordt informatie aan archieven, series dossiers en stukken toegevoegd. En in digitale registraties van papieren archieven worden metadata ingetypt: titels, namen, onderwerpen, trefwoorden, dateringen, classificaties en rubriceringen zijn allemaal voorbeelden van metadata die kunnen worden toegevoegd aan papieren archiefdocumenten of in databases. Ook klappers, indexen en eigentijdse toegangen vormen metadata over archiefstukken, alhoewel ze tegelijkertijd zelf ook archiefstuk kunnen zijn, met hun eigen metadata. En zelfs de fysieke ordening, de volgorde waarin stukken bij elkaar zitten kan informatie geven over de stukken.

Hiervoor was al gesteld dat digitaal archief conceptueel gezien gewoon archief is, en dus ook voor digitaal archief geldt dat er metadata bij hoort over de context, structuur, inhoud en het beheer door de tijd heen. Een groot verschil met papieren archief is dat een deel van de metadata automatisch ingevuld kan worden. Elk digitaal bestand bevat in zichzelf informatie over de datum van aanmaak en de datum van laatste wijziging, die automatisch wordt ingevuld en die bij het kopiëren van het bestand worden mee gekopieerd.

The screenshot displays the 'HUIDIG DOCUMENT' (Current Document) form in the JOIN system. The interface is divided into a left sidebar and a main content area. The sidebar, titled 'Kenniscentrum', contains a tree view of folders: Prodocs, Semi Statistisch, and Trainig feb 2018. Under 'Trainig feb 2018', there are sub-folders for 'Aqualectra', 'A_Incoming', 'A_intern', 'A_Outgoing', 'For Information', 'Governance', and 'HRM'. The 'HRM' folder is expanded, showing 'H_inkomend' (selected), 'H_intern', 'H_uitgaand', 'Job application', 'Payroll', 'PowerPoints', and 'Stocktake'. The main content area shows the document details for 'HUIDIG DOCUMENT'. The document number is 2018-11260, and the scan number is H_inkomend - 22. The input date and time are 22 feb 2018 16:40:50. The sender is a redacted name. The contact person is also redacted. The document date is 22 februari 18. The reference number is empty. The work process is empty. The company is empty. The document type is 'Sollicitatiebrief'. The retention period is 1. The dossier is 'vacature Recordsmanager feb 2018'. The employee ID is empty. The content description is 'sollicitatie vacature Records'. The registration is by 'Me...' (redacted). The 'Afgehandeld' checkbox is unchecked. The date of processing is empty. The global number is 11260. The 'Alleen zichtbaar voor' field is empty. The barcode is empty. A tooltip on the right says 'Sleep bestand(en) hier naar toe of klik om te bladeren'.

Afbeelding: metadata over een document in JOIN. Links de structuur (met wat context), rechts de inhoud (met wat context)

Digitale archiefstukken kunnen worden bewaard met een digitaal informatiesysteem, waarbij software de gegevens over auteur, versiebeheer en andere documentatie over het gevolgde werkproces automatisch kan vastleggen. Een digitaal informatiesysteem dat gebruikt wordt om archivering van documenten mogelijk te maken, zodat ze authentiek, integer, bruikbaar en betrouwbaar blijven, heet een 'archiveringssysteem'. Een digitaal archiveringssysteem dat zich specialiseert in het duurzaam toegankelijk houden van digitale archiefstukken heet ook wel een 'e-depot'. Daarover in een later hoofdstuk meer.

2.1 Context, inhoud en structuur

Kernelementen in de definitie van metadata, papier en digitaal, zijn context, inhoud en structuur.

Context

De context van een archiefstuk zegt iets over hoe de wereld om het archiefstuk heen er uitzag toen het archiefstuk werd gemaakt¹²: wie heeft het stuk gemaakt, onder wiens verantwoordelijkheid en met welk mandaat, binnen welk werkproces, welke wet- en regelgeving was van toepassing, etc. Ook alle informatie over het beheer van het archiefstuk sinds zijn creatie behoren tot de context: wie was er verantwoordelijk voor het archiefbeheer, welke regels en procedures golden er, wat was de ordening of classificatie waaruit de samenhang met andere delen van het archief blijkt, etc.

Inhoud

De inhoudelijke metadata beschrijft waar een archiefstuk, dossier, serie, rubriek of archief over gaat: via een titel, omschrijving, datering en/of trefwoorden wordt een (beknopte) samenvatting gegeven van de inhoud of het onderwerp van een archiefstuk of aggregaties van archiefstukken.

Structuur

De structuur van een archief of een archiefstuk zegt iets over de onderlinge samenhang. Dat kan binnen een document zelf zijn (de structuur van dit document blijkt uit de inhoudsopgave) en tussen documenten onderling: in een dossier, in series samenhangende dossiers, in rubrieken van eenzelfde archiefvormer, en tussen archieven van verschillende archiefvormers.

Binnen een structuur kennen documenten ook een ordening. Documenten kunnen op datum (chronologisch) geordend zijn, of op documenttype (bijvoorbeeld alle notulen of ingekomen post bij elkaar), op naam (alfabetisch) of per zaak (zaaksgewijs) of per werkproces binnen de archiefvormer.

We kennen twee soorten structuur: de logische structuur en de fysieke structuur. Onder de logische structuur verstaan we de inhoudelijke structuur waarin archiefstukken aan de gebruiker getoond worden. De fysieke structuur gaat over de plek en volgorde waarin de fysieke documenten in de kast van een depot staan. De fysieke structuur van een archief kan afwijken van de logische structuur, omdat dozen, planken, kasten en depots een beperkte opslagcapaciteit hebben of omdat sommige typen archiefstukken andere bewaaromstandigheden vereisen. Denk bijvoorbeeld aan foto's, videobanden of landkaarten die logisch gezien bij een bepaald dossier horen, maar die fysiek op een hele andere plek bewaard kunnen worden.

Digitaal zijn de fysieke en logische structuur altijd anders. De fysieke locatie op de harde schijf, usb-stick of cd-rom zijn voor computergebruikers onzichtbaar, en stukjes van een digitaal bestand kunnen in willekeurige volgorde op het opslagmedium staan.

¹² Voor meer informatie over de context van een archiefstuk, zie Hans Hofman, "Een uitdijend heeal?", in : SAP Jaarboek 2000, p. 45-65, <https://kvan.courant.nu/periodicals/JB/2000>

2.2 Technische metadata

Naast metadata over de context, inhoud en structuur kennen digitale archiefstukken ook technische metadata. Hiertoe behoren onder andere het bestandsformaat, de datum van aanmaak en datum van laatste wijziging die door het computersysteem worden bijgehouden, en de omvang van het bestand.

Checksums

Een speciaal onderdeel van technische metadata vormen *checksums*. Hierbij worden cryptografische formules gebruikt om bij een digitaal bestand een controlegetal te berekenen. Deze cryptografische formules hebben twee belangrijke eigenschappen:

1. Wanneer het twee keer over hetzelfde bestand wordt berekend, dan zal het altijd hetzelfde controlegetal als resultaat geven;
2. Wanneer het over twee verschillende bestanden (al zijn ze maar op één bit afwijkend) worden berekend, dan zullen er altijd verschillende controlegetal als resultaat worden gegeven.

Door het berekenen en bewaren van een checksum over een bestand kan vanaf dat moment altijd worden gecontroleerd of het bestand nog identiek is, of dat het is gewijzigd.

2.3 XML

XML¹³ is een manier om gestructureerde informatie op te slaan en uit te wisselen tussen computers. Zowel de inhoud als de structuur zitten in één document. Wanneer de structuurinformatie goed is gekozen, kan een menselijke lezer ook herkennen wat de informatie en de structuur is, zelfs al was de informatie niet voor mensenogen bedoeld.

Deze module is niet bedoeld als hele technische module, maar wie zich nu met metadata over digitale archieven bezighoudt komt zonder twijfel in aanraking met XML. In deze module wordt niet van de studenten verwacht dat ze zelf XML gaan maken. Wel is het de bedoeling dat men vertrouwd raakt met wat XML is, en hoe het gelezen kan worden.

Stel nu, ik heb informatie over een persoon: zijn voornaam, tussenvoegsel en achternaam:

Voornaam:	Jeroen
Tussenvoegsel:	van
Achternaam:	Luin

In XML wordt dit weergegeven als:

```
<voornamen> Jeroen </voornamen>  
<tussenvoegsel> van </tussenvoegsel>  
<achternaam> Luin </achternaam>
```

De aanduiding van de informatie staat dus tussen punthaken, één keer vóór de inhoud, en één keer achter de inhoud, en dan voorzien van extra schuine streep (/). Deze aanduidingen heten *tags*, waarbij de tag vóór de inhoud wordt aangeduid als de open-tag, en de tag na de inhoud de sluit-tag genoemd wordt. Tussen deze twee tags staat de inhoud.

¹³ Voor meer informatie, zie de Wikipedia-pagina: https://nl.wikipedia.org/wiki/Extensible_Markup_Language

Het bovenstaande voorbeeld is vrij simpel, het bevat de naam van één persoon. Maar wat nu wanneer er meerdere personen zijn, hoe hou je dan bij elkaar welke voornaam, tussenvoegsel en achternaam bij elkaar horen?

Eén manier om dat op te lossen is om de personen te groeperen als persoon:

```
<persoon>
  <voornaam> Jeroen </voornaam>
  <tussenvoegsel> van </tussenvoegsel>
  <achternaam> Luin </achternaam>
</persoon>
<persoon>
  <voornaam> Gijs </voornaam>
  <tussenvoegsel></tussenvoegsel>
  <achternaam> Boink </achternaam>
</persoon>
<persoon>
  <voornaam> Maartje </voornaam>
  <tussenvoegsel> van de </tussenvoegsel>
  <achternaam> Kamp </achternaam>
</persoon>
```

De informatie over de personen is zo steeds gegroepeerd. Binnen de openings-tag en sluitings-tag van elke persoon staat de informatie die bij één en dezelfde persoon hoort.

Stel dat we nu bij alle drie personen een personeelsnummer weten, en dat bij Jeroen het personeelsnummer 15 hoort, bij Gijs het nummer 22 en bij Maartje nummer 38. In de XML-structuur kan dit worden opgenomen als toevoeging bij de openings-tag. Dit heet dan een *attribuut*:

```
<persoon nummer="15" >
  <voornaam> Jeroen </voornaam>
  <tussenvoegsel> van </tussenvoegsel>
  <achternaam> Luin </achternaam>
</persoon>
<persoon nummer="22" >
  <voornaam> Gijs </voornaam>
  <tussenvoegsel></tussenvoegsel>
  <achternaam> Boink </achternaam>
</persoon>
<persoon nummer="38" >
  <voornaam> Maartje </voornaam>
  <tussenvoegsel> van de </tussenvoegsel>
  <achternaam> Kamp </achternaam>
</persoon>
```

Voor een computer is het nu heel simpel om antwoord te geven op de vraag "Wat is de voornaam van de persoon met nummer 38?" Maar doordat in deze voorbeelden de tags herkenbare, betekenisvolle namen hebben, is het voor een mens ook leesbaar.

2.4 Oefeningen

Opgave 2.1:

Universiteiten houden van elke student een dossier bij. Schrijf op:

- welke documenten verwacht je in het dossier aan te treffen?
- welke metadata verwacht je op het niveau van het hele dossier over een student?

- c) welke metadata verwacht je voor elk document in een dossier?
- d) hoe zou een papieren dossier bewaard worden, en waar wordt dan de metadata opgeslagen?
- e) hoe zou een digitaal dossier bewaard worden, en waar wordt dan de metadata opgeslagen?

Opgave 2.2:

Op Moodle staan de presentaties die gebruikt zijn bij de lessen van deze opleiding. Download één van de presentaties en kijk welke metadata je over dit bestand kunt vinden. Beantwoord de vragen:

- a) tref je context, inhoud, structuur en technische metadata aan?
- b) welke metadata is automatisch ingevuld door de computer waarmee de presentatie is gemaakt?
- c) welke metadata is handmatig ingevuld door de auteur?
- d) kun je aan de metadata zien waarvandaan je de presentatie hebt gedownload?

Opgave 2.3:

Checksums kunnen worden ingezet om de integriteit van een document te bewaken, doordat je kunt zien of een document al dan niet opzettelijk is gewijzigd. Wat zou je nog meer met een checksum kunnen doen?

Opgave 2.4:

Bekijk het volgende stukje XML, en beschrijf in eigen woorden welke informatie je hier ziet.

```

<Huwelijksakte nummer="9.39v" >
  <Persoon id="1" >
    <Voornamen> Jacobus Franciscus </Voornamen>
    <Achternaam> Vlietman </Achternaam>
    <Geboortedatum>
      <Dag> 15 </Dag>
      <Maand> 4 </Maand>
      <Jaar> 1833 </Jaar>
    </Geboortedatum>
    <Geboorteplaats> Amsterdam </Geboorteplaats>
    <Beroep> Schoenmaker </Beroep>
  </Persoon>
  <Persoon id="2" >
    <Voornamen> Johanna Maria Petronella </Voornamen>
    <Achternaam> Leurink </Achternaam>
    <Geboortedatum>
      <Dag> 28 </Dag>
      <Maand> 5 </Maand>
      <Jaar> 1842 </Jaar>
    </Geboortedatum>
    <Geboorteplaats> Amsterdam </Geboorteplaats>
    <Beroep> zonder </Beroep>
  </Persoon>
  <Huwelijk>
    <Datum>
      <Dag> 1 </Dag>
      <Maand> 6 </Maand>
      <Jaar> 1864 </Jaar>
    </Datum>
    <Bruidegom> 1 </Bruidegom>
    <Bruid> 2 </Bruid>
  </Huwelijk>
</Huwelijksakte>

```

3. Wat is een e-depot?

In de module Archiveringsystemen zijn de begrippen ‘informatiesysteem’ en ‘archiveringssysteem’ aan bod gekomen. Een informatiesysteem werd daarbij voorgesteld als het geheel van bestanden, metadata, procedures, apparaten en daarbij benodigd hulpmiddelen, ingericht door een persoon, groep personen of organisatie ten behoeve van de uitvoering van zijn of haar taken. Onderdeel van een digitaal informatiesysteem is software waarin de documenten worden opgeslagen, voorzien van metadata over:

- Registratie en opslag
- Vastleggen ‘opsteller’
- Toekennen toegangsrechten
- Documenten delen, routeren, advies vragen etc.
- Uitchecken en inchecken, versies maken
- Terugzoeken, etc.

Een archiveringssysteem is een informatiesysteem met extra functies die bedoeld zijn om de authenticiteit van gearchiveerde documenten te garanderen. De definitie die we hanteren is¹⁴:

Het geheel van documenten, metadata, processen, methoden, procedures, kennis, regels, middelen en mensen waarmee een persoon of organisatie zich voorziet van betrouwbare en duurzame informatie ten behoeve van bedrijfsvoering, herinnering en verantwoording.

De extra functies om die betrouwbaarheid en duurzaamheid van informatie te verzorgen zijn:

- Bevriezen van documenten en hele dossiers
- Bijhouden van de audit-trail over alles wat er met een document of dossier gebeurt, door wie dat is gedaan en wanneer dat is gedaan.
- Bijhouden van extra metadata (bijv. bewaartermijn, classificatie, openbaarheid)

Een e-depot is een digitaal archiveringssysteem dat gespecialiseerd in het duurzaam toegankelijk houden van digitale bestanden.

In de praktijk wordt overigens het softwarepakket dat onderdeel is van het archiveringssysteem zelf ook als ‘e-depot’ aangeduid, en spreekt men over “iets in het e-depot stoppen” of “iets uit het e-depot halen”. Hou bij het lezen over e-depots dus altijd rekening met welke betekenis op dat moment aan ‘e-depot’ wordt gegeven: de ruime betekenis zoals Horsman die geeft, of de nauwe betekenis van alleen het softwarepakket.

Voor de noodzaak om een e-depot (in ruime zin) te hebben zijn vier redenen:

1. Digitale opslagmedia kunnen verouderd raken

Floppy disks, diskettes, zip-disks en cd-roms zijn in veel moderne computersystemen al niet meer te lezen. Voor duurzame toegang tot de informatie die op die opslagmedia staat is het nodig deze naar moderne opslagmedia te verplaatsen. En wanneer die moderne opslagmedia zelf weer verouderd zijn, de informatie weer verder te verhuizen;

¹⁴ Uit: Peter Horsman, “Archiveren, een inleiding”, 2011, <https://kvan.courant.nu/issue/AS/2011-01-01/edition/0/page/1>

2. **De bits en bytes van een bestand kunnen beschadigd raken**

Elk digitaal bestand bestaat uit bits en bytes (de 'bitstream') die op een opslagmedium zijn opgeslagen en via een netwerk kunnen worden getransporteerd. Tijdens de opslag en tijdens het transport kunnen deze bitstreams beschadigd raken, waardoor informatie verloren gaat. Door meerdere kopieën van een bitstream op meerdere locaties te bewaren en deze regelmatig met elkaar te vergelijken (bijvoorbeeld middels de eerder genoemde *checksum*) kunnen fouten snel worden opgespoord en kan de foutieve kopie worden vervangen;

3. **Bestandsformaten kunnen verouderd raken**

De levensduur van hardware en software is, vergeleken met papier, heel kort. Waar het in de jaren '90 van de vorige eeuw nog heel normaal was om documenten in WordPerfect, spreadsheets in Lotus 123 en databases in DBase 3 te maken, zijn deze bestandsformaten tegenwoordig niet meer te gebruiken. En de bestandsformaten die we tegenwoordig gebruiken zullen op een dag ook vervangen worden door iets nieuws. Een e-depot moet in staat zijn om informatie in bestanden met een verouderd bestandsformaat toegankelijk te houden, bijvoorbeeld door omzetting naar een moderner bestandsformaat (migratie), of door op moderne computers de oude hardware en software na te bootsen (emulatie);

4. **De gebruiker verandert**

Niet alleen de hardware en software worden vernieuwd, ook de gebruikers van het archief veranderen. Terminologie die nu gebruikt wordt, zal door het publiek van over 100 jaar wellicht niet meer gesnapt worden, net zoals wij nu moeite kunnen hebben met het lezen of begrijpen van teksten die 100 jaar geleden geschreven zijn.

3.1 Functies van een digitaal archiveringssysteem

In het eerste deel van dit hoofdstuk werd een aantal functies aangegeven voor een archiveringssysteem. In de NEN-ISO 15489-1:2016 norm voor informatie- en archiefmanagement wordt een gestandaardiseerde lijst processen genoemd die met behulp van een archiveringssysteem moeten kunnen worden uitgevoerd:

1. **Aanmaken van archiefstukken**

Dit gebeurt over het algemeen buiten het archiefsysteem, maar valt wel binnen de ruime definitie van een archiveringssysteem: het geheel van mensen, middelen, etc. Bij het aanmaken horen zowel de inhoud als de metadata te worden vastgelegd;

2. **Opnemen van archiefstukken in een archiefsysteem**

Wanneer uit het proces van waardering blijkt dat het nodig is archiefstukken gedurende langere tijd te behouden en te beheren, moeten deze samen met de bijhorende metadata in een archiefsysteem worden opgenomen. Hierbij moeten ze worden voorzien van een uniek identificatiekenmerk, moet informatie over het proces van opname worden vastgelegd en moeten relaties met actoren, werkprocessen en/of andere archiefstukken worden aangebracht;

3. **Classificatie en indexering van archiefstukken**

Het gebruikte archiveringssysteem moet voorzien zijn van een classificatieschema dat de archiefstukken koppelt aan de bedrijfscontext door middel van categorieën die horen bij de verschillende bedrijfsactiviteiten. In het archiveringssysteem kan daarnaast ook met indexering worden gewerkt, waarbij trefwoorden voor bijvoorbeeld onderwerpen, locaties of personen worden toegekend;

4. Toegangscontrole tot de archiefstukken

De toegang tot de archiefstukken moet via autorisaties worden beheerd, zodat alleen geautoriseerde personen de archiefstukken kunnen beheren of raadplegen. Autorisaties moeten kunnen worden toegekend, veranderd en ingetrokken door daartoe bevoegde personen, en deze wijzigingen moeten zelf ook weer worden vastgelegd;

5. Opslaan van archiefstukken

Archiefstukken moeten in het archiveringsysteem zodanig worden opgeslagen dat ze beschermd zijn tegen ongeautoriseerde toegang, verandering, verlies of vernietiging, met inbegrip van diefstal en rampen. Hiertoe moet gekozen worden voor geschikte opslag-omgevingen die goed beheerd worden, routinematig beschermd worden en voorzien zijn van goede rampenplannen en geteste herstelprocedures;

6. Gebruik en hergebruik van archiefstukken

Archiefstukken moeten zolang ze bewaard worden ook vindbaar en bruikbaar zijn. De archiefstukken moeten kunnen worden geraadpleegd of worden geëxporteerd gebruikmakend van alle opgenomen metadata en classificaties, voor een ieder die daartoe de juiste autorisatie heeft. Hiertoe kan het nodig zijn om metadata te wijzigen of raadpleeg-versies van een archiefstuk te maken;

7. Migratie en conversie

Voor de duurzame toegankelijkheid tot archiefstukken kan het nodig zijn ze middels migratie (overzetten naar een nieuwe versie van hetzelfde bestandsformaat) of conversie (overzetten naar een heel nieuw bestandsformaat) om te zetten naar een nieuw bestand. De processen van migratie en conversie moeten beheerd en gedocumenteerd worden uitgevoerd;

8. Verwijdering

Archiefstukken die niet langer bewaard hoeven te worden, bijvoorbeeld omdat de bewaartermijn verlopen is of omdat het betreffende archief overgedragen of afgestoten wordt, moeten volgens een beheerd proces en met de juiste autorisaties worden verwijderd.

3.2 Kenmerken van digitale archiveringsystemen

Volgens de NEN-ISO 15489-1:2016 moeten archiveringsystemen, en dus ook digitale archiveringsystemen aan de volgende kwaliteitskenmerken voldoen:

1. Betrouwbaar

Een archiveringsysteem is betrouwbaar wanneer het continu en regelmatig naar behoren alle in de vorige paragraaf genoemde functies uitvoert;

2. Veilig en integer

Er moeten voldoende maatregelen aanwezig zijn om ongeautoriseerde toegang, wijziging, verberging of vernietiging van archiefstukken te voorkomen. Alle wel geautoriseerde toegang, wijziging, verberging en vernietiging wordt gedocumenteerd;

3. Afdoende en compliant

Het beheer van de archiefstukken in het archiveringsysteem voldoet aan de verwachtingen van de gebruikers van het systeem en aan de geldende wet- en regelgeving. Compliancy met deze verwachtingen en regels wordt regelmatig gecontroleerd;

4. Uitgebreid

Alle soorten archiefstukken uit de volledige verzameling van bedrijfsactiviteiten van de organisatie moeten kunnen worden beheerd;

5. Systematisch

Het aanmaken, opnemen en beheren van archiefstukken vindt systematisch plaats (en dus niet ad-hoc) door de routinematige werking en door het volgen van opgesteld beleid en procedures.

3.3 OAIS

Voor digitale archiveringssystemen is een verzamelnaam bedacht: *Open Archival Information System* of OAIS. Onder dezelfde naam en met dezelfde afkorting is ook een ISO referentiemodel gemaakt waarmee eisen aan goede e-depots worden beschreven: ISO 14721.

Het OAIS-model beschrijft onder andere:

- De verantwoordelijkheden van een OAIS
- Een functioneel model voor de functies waaraan een e-depot moet voldoen;
- Een begrippenlijst voor veel voorkomende onderdelen van een e-depot;
- De omgeving van het e-depot: de archiefvormers, het (beleids-)management van het e-depot en de gebruikers van het e-depot.

Verantwoordelijkheden van een OAIS

Volgens het OAIS referentiemodel heeft een OAIS archief de volgende verantwoordelijkheden:

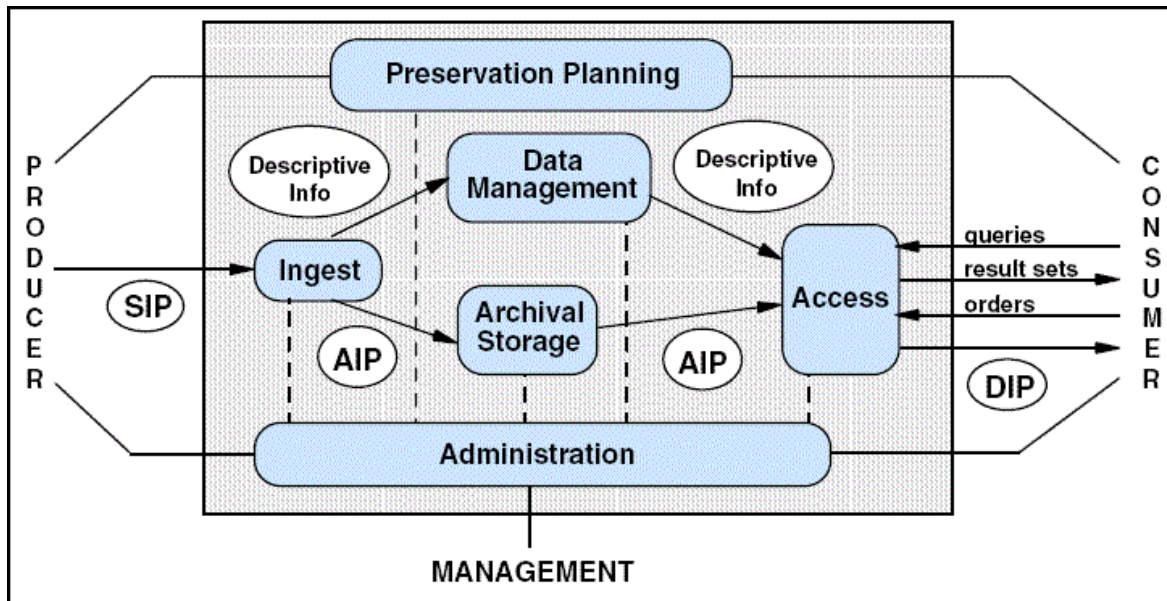
- Zorgen voor goede afspraken met de archiefvormers zodat alle benodigde informatie wordt aangeleverd bij een overbrenging of uitplaatsing.
- Zorgen voor voldoende (juridisch) mandaat om het beheer te kunnen en mogen uitvoeren.
- Bepalen van de beoogde gebruikersgroep, of Designated Community (zie hierna bij 'Begrippen')
- Zorgen voor voldoende toegankelijkheid van de informatie in het OAIS-archief
- Bewaren en kunnen aantonen van de authenticiteit van stukken in het OAIS-archief
- Hebben van een duidelijk beleid met goed beschreven procedures en processen

Functioneel model

Het OAIS referentiemodel beschrijft zes bedrijfsfuncties waaraan een e-depot moet voldoen:

1. Ingest, het proces van opname van nieuwe archiefstukken en bijhorende metadata;
2. Data Management, het beheer van de beschrijvende metadata en van de beheerhandelingen die tijdens het beheer op het archief plaatsvinden;
3. Archival Storage, de duurzame opslag en het beheer van de digitale archiefbestanden zelf;
4. Administration, het dagelijks beheer van het e-depot, waaronder autorisaties;
5. Preservation planning, het geheel van systemen, procedures, hulpmiddelen en mensen bedoeld om de duurzame toegankelijkheid van de opgeslagen bestanden en metadata te realiseren;
6. Access, het geven van toegang tot opgenomen archiefbestanden aan geautoriseerde gebruikers

In de afbeelding hieronder wordt het functionele model van het OAIS-referentiemodel schematisch weergegeven.



Afbeelding: het functionele model uit het OAIS-Referentiemodel

OAIS-functies en ISO15489-functies

De twee normen OAIS en ISO15489 geven beide een lijst met functies waaraan een (digitaal) archiveringssysteem moet voldoen, maar ieder op zijn eigen manier. De twee sets functies kunnen naast elkaar worden gezet:

OAIS	ISO 15489-1:2016
Ingest	Opnemen van archiefstukken Classificatie en indexering
Data Management	Classificatie en indexering Toegangscontrole Gebruik en hergebruik Verwijdering
Archival Storage	Opslaan van archiefstukken Migratie en conversie Verwijdering
Preservation Planning	Gebruik en hergebruik Migratie en conversie
Access	Toegangscontrole Gebruik en hergebruik
Administration	Classificatie en indexering Toegangscontrole

Alleen de functie 'Aanmaken van archiefstukken' valt binnen OAIS buiten scope, omdat de archiefvormer wordt gezien als onderdeel van de omgeving van een e-depot, en niet als onderdeel van een e-depot zelf.

Begrippen

In het OAIS-model worden een aantal belangrijke begrippen uitgewerkt die bij digitaal archivering veelvuldig worden gebruikt¹⁵:

- **Producer**
De producent van de archiefstukken en metadata die opgenomen gaat worden in het e-depot.
- **Management**
De beleidsmakers die de randvoorwaarden van het e-depot realiseren, die strategische en tactische beslissingen over het e-depot nemen, die voor de financiering en voor het mandaat zorgen, en die bepalen wat de missie en doelstellingen van het e-depot zijn, etc.
- **Consumer**
De gebruiker van archiefmateriaal dat in het e-depot is opgenomen. Onderdeel van de *designated community* (zie hieronder).
- **Submission Information Package (SIP)**
Het geheel van archiefbestanden en bijhorende metadata dat door een archiefvormer wordt aangeleverd om te worden opgenomen in een e-depot;
- **Archival Information Package (AIP)**
Het geheel van archiefbestanden en bijhorende metadata dat na opname in het systeem is opgeslagen en wordt beheerd;
- **Dissemination Information Package (DIP)**
Het geheel van archiefbestanden en bijhorende metadata dat aan een archiefgebruiker ter beschikking wordt gesteld;
- **Representation information**
Informatie die los staat van een informatieobject zelf, maar die gebruik kan of moet worden om de data in het informatieobject te interpreteren. Uitleg over de archiefvormer in de inleiding van een inventaris kan representatie-informatie zijn, maar ook uitleg over met welke software een specifiek bestandsformaat geopend kan worden of een lijstje met uitleg over Latijnse termen die in een document kunnen worden gebruikt;
- **Designated Community**
De groep(en) beoogde gebruikers van de informatie die in het e-depot bewaard wordt, en die in staat zouden moeten zijn om gevonden archiefstukken met behulp van de metadata en de representatie-informatie te kunnen interpreteren.
- **Knowledge Base**
Van de gebruikers uit de designated community wordt enige voorkennis verwacht, bijvoorbeeld het kunnen lezen, de taal van de archiefstukken kennen, oude handschriften kunnen lezen of kennis van de staatkundige indeling van een land. Deze knowledge base kan door de tijd heen wijzigen. Dit moet door de archiefinstelling die een e-depot beheert goed in de gaten gehouden worden. Wanneer de knowledge base zoveel veranderd is dat de gebruikers de stukken niet meer kunnen raadplegen, zal extra representatie-informatie moeten worden toegevoegd in de vorm van extra uitleg of onderzoeksgidsen.

¹⁵ Voor een uitgebreidere beschrijving van deze begrippen, zie het artikel van Barbara Sierman, "Het OAIS-model, een leidraad voor duurzame toegankelijkheid", [IWA-base nummer 5767-1](#).

3.4 Oefeningen

Opgave 3.1:

In het begin van dit hoofdstuk zijn 4 redenen aangegeven die de duurzame toegankelijkheid van digitale archiefstukken bedreigen: verouderende informatiedragers, kapotte bitstreams, onleesbare bestandsformaten en een veranderend publiek.

- a) Geef van elk van de vier redenen een voorbeeld waar je in jouw eigen werkomgeving tegenaan bent gelopen, of waar je tegenaan zou kunnen lopen.
- b) Voor de voorbeelden uit a) waar je in de praktijk al tegenaan gelopen bent, hoe hebben jullie dat opgelost? Voor de voorbeelden die zich in de praktijk nog niet hebben voorgedaan: hoe zou je dat nu op gaan lossen?

Opgave 3.2:

De NEN-ISO 15489 standaard kent 8 functies die elk archiveringssysteem idealiter zou moeten ondersteunen (zie pagina 15). Zijn al deze 8 functies noodzakelijk om een goed archiveringssysteem te zijn? Welke functies zouden weggelaten kunnen worden zonder dat dit de kwaliteit van het archiveringssysteem aantast?

Opgave 3.3:

De OAIS-norm is opgenomen in een ISO-standaard. Als archiefdienst kun je jezelf laten certificeren als 'OAIS-compliant'¹⁶, waarbij je aantoonbaar alle taken en verantwoordelijkheden van een OAIS archief uitvoert. Hierbij wordt dus niet alleen gekeken naar het softwarepakket (e-depot in 'nauwe' zin) maar wordt gekeken naar de hele organisatie (e-depot in 'ruime' zin) met alle processen, procedures, mensen, middelen, etc. Waarom kan alleen een softwarepakket op zichzelf niet OAIS-compliant zijn?

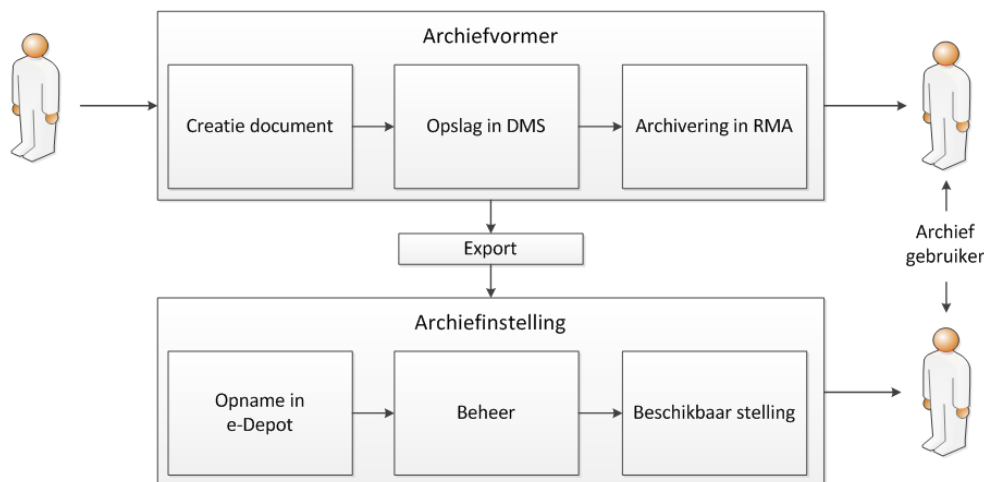
Opgave 3.4:

Geef een voorbeeld van digitale informatie waar jij in je dagelijks werk mee omgaat, waar een buitenstaander zonder extra uitleg niets van zou snappen. Welke extra informatie ('representation information') zou je moeten meeleveren wil deze buitenstaander het wel snappen? Is die informatie bij jullie aanwezig en wordt die informatie mee-gearchiveerd?

¹⁶ Zie bijvoorbeeld <https://www.coretrustseal.org/>

4. Proces van digitale archivering

Het proces van digitale archivering van creatie tot aan beschikbaarstelling kan schematisch als volgt worden weergegeven:



In dit proces vindt dus op twee momenten een overdracht plaats. Als eerste het moment van overzetting van het primaire proces naar de eigen archiefafdeling van de archiefvormer. Hierbij is de eigen organisatie de grootste klant, maar kan er ook voor derden onderzoek in het archief worden gedaan. Later kunnen alle blijvend te bewaren archiefstukken worden overgebracht naar een archiefinstelling, of kan er besloten worden om een deel van het archief uit te plaatsen in het e-depot van een archiefinstelling. Waar in dit hoofdstuk wordt gesproken over 'de archiefdienst' kan zowel de archiefafdeling binnen de archiefvormer, als een archiefinstelling worden gelezen.

De overdracht van archiefstukken vanuit het bronsysteem naar een e-depot is een kritiek moment: na opname moet de archiefvormer de overgedragen bestanden en metadata uit het bronsysteem verwijderen, ze zijn nu immers overgebracht en de verantwoordelijkheid van de archiefdienst geworden. Dat betekent ook dat een overbrenging later niet overnieuw gedaan kan worden wanneer er iets fout blijkt te zijn gegaan. Zit het eenmaal goed in het e-depot kun je het beheer en de beschikbaar stelling altijd aanpassen en verbeteren, maar het overbrengen en opnemen in het e-depot moet foutloos worden uitgevoerd¹⁷.

Om de overdracht van de digitale archiefdocumenten naar een e-depot mogelijk te maken moet het bronsysteem van de archiefvormer worden aangesloten op het e-depot. Zo'n aansluiting vereist organisatorische, inhoudelijke en technische afstemming. Voor het aansluiten is door het Nationaal Archief in Nederland het instrument 'Impactanalyse aansluitingen' ontwikkeld.

4.1 Impactanalyse aansluitingen

Met een impactanalyse onderzoekt een werkgroep met mensen van de archiefdienst en van de archiefvormer, of het op te nemen archief voldoet aan de eisen die de archiefdienst stelt aan

¹⁷ Enig realisme is hier wel op z'n plaats: helemaal foutloos zal het nooit kunnen zijn. We streven naar een goed gecontroleerd mogelijk proces om de kans op fouten zo klein mogelijk te houden.

archiefmateriaal. Zo niet, dan wordt er gekeken wat er nog moet gebeuren om uitplaatsen of overbrengen mogelijk te maken. Hierbij wordt geen onderscheid gemaakt tussen born-digital archiefmateriaal en scans die vervanging zijn van papieren archief.

Door inschattingen te maken beantwoordt de werkgroep de vragen:

- Wat moet er gebeuren om het archiefsysteem aan te sluiten op het e-depot?
- In welke volgorde moet dat gebeuren?
- Welke randvoorwaarden zijn van toepassing?
- Welke mensen (rollen/funcities/kennis) zijn hierbij nodig?
- Hoeveel tijd is hiervoor nodig?
- Wat zijn de kosten voor de archiefvormer en de archiefdienst?

Het doel van de impactanalyse is om een projectplan te hebben waarmee zowel de archiefvormer als de archiefdienst weten:

7. Wat de archiefvormer moet doen om het digitale archief in goede, geordende en toegankelijke staat aan te leveren
8. Wat de archiefvormer en de archiefdienst moeten doen om het digitale archief in het e-depot op te nemen, te beheren, en beschikbaar te stellen
9. Wat de archiefvormer en de archiefdienst moeten doen om de koppeling te realiseren voor één van de volgende transportmethoden:
 - a. een permanente en duurzame verbinding op basis van de Nederlandse overheidsstandaard Digikoppeling;
 - b. een beveiligde FTP verbinding;
 - c. een losse mediadrager.
10. Welke juridische en organisatorische maatregelen nodig zijn om aan te sluiten op een e-depot.

Om zeker te weten dat de aansluiting niet alleen in theorie maar ook echt in de praktijk werkt, wordt na afloop van de impactanalyse het aansluittraject één keer helemaal doorlopen door echt een overbrenging of uitplaatsing te doen. Dat mag best een klein archiefblok zijn, zolang alle stappen van de aansluiting en de uitplaatsing of overbrenging maar daadwerkelijk worden doorlopen.

Proces van impactanalyse

Het proces van de impactanalyse bestaat uit 6 stappen:

1. **Samenstellen van de werkgroep**
Hierin zitten experts van de archiefdienst met betrekking tot digitale archivering en het e-depot, inhoudelijke experts van de archiefvormer en experts van informatiesysteem waarin de documenten nu zijn opgeslagen.
2. **Ontvangst van een representatieve dataset**
De inhoudelijke en de technische analyse wordt uitgevoerd op een stukje van het archief dat aan het einde van de impactanalyse uitgeplaatst of overgebracht gaat worden. Deze dataset moet een afspiegeling zijn van het type documenten en de soort ordening die in de volledige overbrenging of uitplaatsing gaat voorkomen.

3. **Kick-off / startbijeenkomst**

Tijdens de kick-off stellen de werkgroepleden zichzelf aan elkaar voor, wordt het doel en het proces van de impactanalyse uitgelegd, en geven de experts van de archiefvormer uitleg over hoe het bronsysteem er uitziet en wat er in de representatieve dataset zit. Aan het einde van de startbijeenkomst worden de aanwezigen in 3 teams opgesplitst. Eén team onderzoekt de organisatorische vragen, één team de inhoudelijke vragen en één team de technische vragen.

4. **Onderzoek en analyse**

In één of twee weken na de kick-off doen de teams onderzoek en proberen de vragen te beantwoorden.

5. **Vervolgbijeenkomst**

Na één of twee weken onderzoek doen de teams verslag aan elkaar over wat ze aan antwoorden gevonden hebben. De overige werkgroepleden kunnen vragen stellen over verduidelijking of gevonden obstakels proberen op te lossen

6. **Schrijven eindrapport**

Elk team doet na afloop van de tweede bijeenkomst eventueel nog wat vervolgonderzoek, en schrijft vervolgens het eindrapport van het team. Tezamen vormen de drie teamrapporten een compleet overzicht van alle stappen die voor een uitplaatsing of overbrenging nodig zijn.

Na afloop van de impactanalyse is bekend welke stappen in een aansluitproject moeten worden uitgevoerd om de aansluiting daadwerkelijk uit te voeren. Een projectleider van de archiefvormer en een projectleider van de archiefdienst voeren samen dit project uit. In dit aansluitproject wordt daadwerkelijk een uitplaatsing of overbrenging uitgevoerd, volgens het hieronder genoemde proces.

4.2 **Procesbeschrijving overbrenging of uitplaatsing**

Het proces van overbrenging of uitplaatsing wordt gestart door een verzoek van een archiefvormer aan de archiefdienst om een archiefblok over te brengen of uit te plaatsen. De medewerker van de archiefdienst die de verwerving gaat uitvoeren begint met het aanmaken van een document waarin de afspraken met de archiefvormer over de overbrenging of uitplaatsing worden vastgelegd: het beslisdocument. Wanneer de archiefdienst en de archiefvormer niet zelf de tijdens de impactanalyse gevonden problemen willen of kunnen oplossen, dan kunnen ze een derde partij als archiefbewerker inschakelen. Deze archiefbewerker schrijft dan mee aan het beslisdocument en ondertekent ook mee.

Beslisdocument

Het beslisdocument is dus een lijst met afspraken tussen de archiefvormer, de archiefdienst en een eventuele archiefbewerker en wordt gevuld met de informatie die tijdens de impactanalyse is opgedaan en met het resultaat van de metadata-mapping. Over de metadata-mapping volgt in een volgende paragraaf meer.

Een beslisdocument bestaat uit de volgende onderdelen:

1. **Voorblad** met daarop een samenvatting van de administratieve gegevens (wie, wat, wanneer) en informatie over versiebeheer van het document.
2. **Contactgegevens en projectrollen**, met de namen en contactgegevens van de aanspreekpunten vanuit de archiefvormer, de archiefdienst en eventueel de archiefbewerker, en informatie over (de geschiedenis van) de archiefvormer.

3. **Informatie over het archiefblok** dat overgebracht of uitgeplaatst gaat worden, waaronder een samenvatting van de inhoud, verwante archieven (bij de archiefvormer of bij de archiefdienst), de locatie, omvang en bestandsformaten van de bestanden in het archiefblok en of er openbaarheidsbeperkingen¹⁸ gesteld moeten worden.
4. **Waardering, selectie en vernietiging**, waarin wordt aangegeven hoe de stukken in het archiefblok gewaardeerd zijn en op welk niveau (rubriek, subrubriek, dossier, stuk) de waardering is aangebracht, of er uitzonderingen zijn waarbij stukken die eigenlijk vernietigd zouden worden toch als ‘Bewaren’ worden aangemerkt, en wie de stukken die vernietigd moeten worden uiteindelijk gaat vernietigen
5. **Toekennen metadata**, waarbij wordt aangegeven in welke vorm de metadata beschikbaar zijn, en of ze al voldoen aan het metadataschema dat door het Nationaal Archief wordt voorgeschreven. Ook kan hier worden aangegeven of er extra metadata (bijlagen, concordanties¹⁹ of eigentijdse toegangen²⁰) beschikbaar zijn.
6. **Archiefbewerkingsplan**, waarin wordt beschreven welke inhoudelijke en technische bewerkingen er nog op het archiefblok moeten worden uitgevoerd voordat het in goede, geordende en toegankelijke staat is. Deze informatie die hier moet worden ingevuld is tijdens de impactanalyse grotendeels opgesteld, zoals wat er in de representatieve dataset is aangetroffen, welke afwijkingen ten opzichte van de geldende normen²¹ is aangetroffen, welke optionele extra bewerkingen er afgesproken worden en welke bewerkingen er nog in de toekomst, wanneer het in beheer is van de archiefdienst, moeten worden uitgevoerd.
7. **Transport**, waarin is vastgelegd op welke (technische) manier de transport van de digitale archiefstukken gaat plaatsvinden (bijvoorbeeld via internet, of via een USB-stick of andere losse gegevensdrager), door wie het transport wordt uitgevoerd en wat er dan in het transportpakket zit.
8. **Ondertekening**, door de twee of drie partijen die bij het opstellen van het beslisdocument betrokken zijn geweest. Vanaf het moment van ondertekening liggen de afspraken vast en kunnen de partijen aan de slag om de overbrenging of uitplaatsing te realiseren. Vanaf dat moment mag de archiefdienst ervan uitgaan dat het archiefblok volgens de afgesproken manier wordt aangeleverd, en mag de archiefvormer ervan uitgaan dat, wanneer hij het archiefblok volgens de afspraken heeft bewerkt, het ook door de archiefdienst zal worden geaccepteerd.

Door de snelle technologische veranderingen wordt er een houdbaarheidsdatum op het beslisdocument gezet: geldig tot 6 maanden na de datum waarop de laatste ondertekenaar heeft getekend. Wanneer de 6 maanden verstreken zijn moet er opnieuw gekeken worden naar de afspraken, moeten ze waar nodig worden aangepast, en opnieuw worden getekend.

9. **Bijlagen: concept archiefordening en metadata-mapping**. De concept-archiefordening met een overzicht van de rubrieken die in het archiefblok zitten, en de metadata-mapping die is uitgevoerd (daarover hierna meer) worden als bijlage aan het beslisdocument toegevoegd.

¹⁸ Archiefwet: artikel 15; Archieflandsverordening: artikel 20

¹⁹ Een tabel waarin na de henummering van een archief staat welk nieuw nummer bij elk oude nummer hoort.

²⁰ Eigentijdse toegangen zijn de toegangen die tijdens de vorming en het gebruik van het archief door de archiefvormer zelf zijn gebruikt.

²¹ Welke normen dat zijn verschilt per archiefdienst. Het Nationaal Archief in Nederland hanteert het [Toepassingsprofiel Metagegevens Rijk](#), de Normen goede en geordende staat, het [Normblad Archiefinventaris](#), de [ISAD\(G\) beschrijvingsstandaard](#), de [Archiefwet 1995](#) en de [lijst voorkeursformaten voor digitale bestanden](#).

Van het beslisdocument is een template beschikbaar dat kan worden ingevuld, en er is een invulinstructie waarbij per in te vullen onderdeel een uitleg is gegeven. Beide zijn te vinden op Moodle.

Metadata-mapping

Eén van de meest tijdrovende onderdelen van een aansluiting is het maken van een metadata-mapping. Het metadataschema van het informatiesysteem waarin een over te brengen of uit te plaatsen archief zit, zal zelden helemaal identiek zijn aan het metadataschema van de archiefdienst. Waar het ene systeem 'auteur' heeft staan, zal een ander 'opsteller' of 'aanmaker' of 'afzender' hebben staan. En 'naam', 'titel', 'onderwerp' of 'beschrijving' worden ook vaak door elkaar gebruikt.

Bij een metadata-mapping gaan experts van het bronsysteem en experts van de archiefdienst samen vastleggen welke metadata-velden van de archiefdienst gevuld gaan worden en waar die informatie uit het bronsysteem gehaald wordt.

ToPX

Voor het maken van metadataschema's is een ISO-standaard ontwikkeld, de ISO 23081. Voor de Nederlandse overheid is die uitgewerkt in een Richtlijn Metagegevens Overheidsinformatie. In deze richtlijn staan de metadata-velden ('elementen') die volgens de Nederlandse overheid ingevuld moeten ('Verplicht' en 'Verplicht indien van toepassing') of mogen ('Aanbevolen' en 'Optioneel') worden. Het is vervolgens aan de overheden zelf om die richtlijn toe te passen op hun eigen informatiesystemen. Hierbij mogen ze er voor kiezen om metadata-elementen die in de Richtlijn als 'aanbevolen' of 'optioneel' zijn aangemerkt alsnog verplicht te stellen.

Voor de Nederlandse Rijksoverheid en voor de Nederlandse lokale overheden zijn twee toepassingsprofielen gemaakt: het Toepassingsprofiel Metagegevens Rijksoverheid (TP-Rijk) en het Toepassingsprofiel Metagegevens Lokale Overheden (TMLO). Beide toepassingsprofielen bevatten dezelfde metadata-elementen, maar verschillen in welke elementen verplicht zijn. Ook is de toelichting met daarin hoe bepaalde velden gevuld moeten worden anders.

Omdat de Nederlandse Rijksoverheid en Provinciale overheden samen hetzelfde e-depot gebruiken is ervoor gekozen om één technische uitwerking van zowel TP-Rijk als TMLO te maken: ToPX, een XML-formaat waarin per metadata-element steeds de lichtste van de verplichtingen uit TP-Rijk en TMLO is gekozen. Een voorbeeld van hoe zo'n XML-structuur eruit ziet staat hieronder bij de Exportvoorwaarden.

Voorbeelden van deze metadata-velden zijn²²:

- **Identificatiekenmerk**, een verplichte unieke code waaronder een archief, rubriek²³, dossier, record of bestand te vinden is. De uniciteit geldt voor dossiers en records binnen de context van het archief. Twee verschillende archieven mogen wel ieder een dossier nummer 1 hebben, maar in één archief mag dossiernummer 1 maar één keer voorkomen.
- **Aggregatieniveau**, een verplichte aanduiding die aangeeft of de metadata bij een archief, rubriek, dossier, record of bestand hoort.
- **Naam**, een verplichte beknopte inhoudelijke beschrijving van het archief, de rubriek, het dossier, het record of het bestand.
- **Classificatie**, een optioneel element waarin de codering van een ordeningsschema kan worden opgenomen, bijvoorbeeld de Basisarchiefcode (BAC).

²² Voor meer uitleg, zie <https://www.nationaalarchief.nl/archiveren/kennisbank/metadata-en-het-e-depot>

²³ De Engelse benaming voor rubriek is 'series', wat in ToPX vertaald is naar 'serie'. Dit is wat verwarrend, omdat in de Nederlandstalige archieftermen 'serie' gereserveerd is voor een aantal gelijkvormige dossiers.

- **Omschrijving**, een optioneel vrij tekstveld waarin uitgebreider dan in de naam de inhoud kan worden opgenomen.
- **Plaats**, een optioneel element waarin de locatie van een map of bestand kan worden opgenomen.
- **Extern identificatiekenmerk**, een optioneel element waarin identificatiekenmerken uit eerdere bronsystemen kunnen worden opgenomen. Wanneer een archiefstuk voorafgaand aan opname in het huidige bronsysteem eerst nog in een ouder systeem heeft gezeten en daar een nummer had, dan kan dat hier worden meegenomen. Mocht een onderzoeker over het oude nummer beschikken, of mocht er vanuit een document verwezen worden naar zo'n oud nummer, dan kan dat alsnog gevonden worden.
- **Eventgeschiedenis**, een optioneel element waarin de beheerhandelingen staan beschreven die het archief, rubriek, dossier, record of bestand heeft ondergaan in het bronsysteem. Hieronder vallen bijvoorbeeld de aanmaak van een dossier, het afsluiten van een dossier, en het registreren van een nieuwe versie van een bestand.
- **Eventplan**, een optioneel element waarin toekomstige beheerhandelingen alvast opgenomen kunnen worden. Bijvoorbeeld het openbaar worden van een beperkt openbaar dossier, of het vernietigen van een uitgeplaatst dossier dat niet blijvend bewaard hoeft te worden.
- **Relaties**, waarin optioneel onderlinge relaties tussen documenten, record of dossiers kunnen worden aangebracht. Dit kan bijvoorbeeld worden gebruikt wanneer twee dossiers over dezelfde persoon of zaak gaan.
- **Actor**, een verplicht element waarin de organisatie of functionaris die verantwoordelijk was voor het maken of opnemen van het archiefstuk, het record, het dossier of het archief is opgenomen. In Nederland wordt hier een verwijzing opgenomen naar het Actorenregister²⁴.
- **Vertrouwelijkheid** (bij uitplaatsing) of **Openbaarheid** (bij overbrenging) waarin staat of er beperkingen zijn aan wie de onderliggende stukken mag raadplegen.
- **Technische metadata**, alleen voor bestanden een set verplichte elementen, zoals bestandsformaat, bestandsnaam, aanmaakdatum, omvang en checksum.

Wanneer een ToPX-element optioneel is, kun je als archiefdienst besluiten dat je voor jezelf het element wel als verplicht beschouwd, zeker wanneer het voor jou geldende toepassingsprofiel (TP-Rijk voor Rijksoverheid of TMLO voor lokale overheden) dat element verplicht stelt.

Niet alle informatie die in ToPX ingevuld moet worden zal letterlijk zo in het bronsysteem te vinden zijn. Soms moeten velden worden berekend of ergens buiten het bronsysteem vandaan worden gehaald, (bijvoorbeeld de verwijzing naar het Actorenregister) of wordt ergens een vaste waarde ingevuld (bijvoorbeeld de titel van het archiefblok). Nadat de mapping voor alle aggregatieniveaus (archief, rubrieken, dossiers, records) is uitgevoerd kan de archiefvormer aan de slag gaan om een exporteer-module bij zijn informatiesysteem te (laten) maken die het gewenste archiefblok met de juiste metadata exporteert.

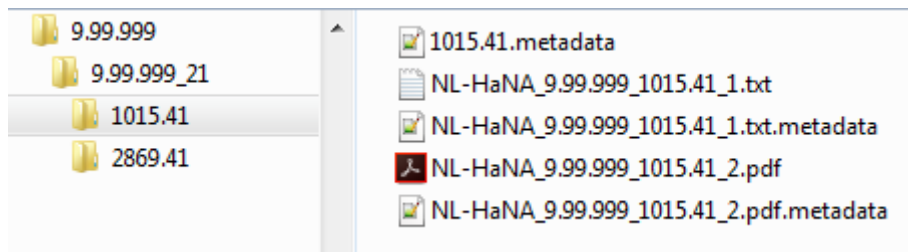
Exportvoorwaarden

Een e-depot bij een archiefdienst, stelt eisen aan hoe de aangeleverde set bestanden met metadata (de SIP) eruit ziet. Welke eisen dat zijn, verschilt per softwarepakket dat onderdeel is van het e-depot.

Het Nederlandse Nationaal Archief gebruikt het softwarepakket Preservica in hun e-depot-omgeving. De SIP die daar moet worden aangeleverd gebruikt een hiërarchische *sidecar*-structuur²⁵. Dit is een

²⁴ Bijvoorbeeld, voor de actor 'Ministerie van Binnenlandse Zaken en Koninkrijksrelaties': <https://hdl.handle.net/10648/79ea38ff-57f1-4893-ad77-23cb6b3b3110>

mappenstructuur zoals je die op de harde schijf ziet met als hoogste niveau het archiefniveau, daarbinnen de series, daarbinnen de dossiers, en daarbinnen de records met bestanden. In elke map zit een tekstbestandje waarin de metadata van dat niveau zit.



Afbeelding: mappenstructuur van archief met code 9.99.999

In bovenstaande afbeelding is de mappenstructuur van een klein archiefje uitgewerkt. Dit archiefje heeft het toegangsnummer 9.99.999, één serie, met code 9.99.999_21, en daarin twee dossiers met nummers 1015.41 en 2869.41. Rechts staat de inhoud van dossier 1015.41. In het dossier zitten twee bestanden (een tekstbestand en een Pdf-bestand) en drie metadatabestanden. Het bovenste metadatabestand heeft dezelfde naam als het dossier met erachter “.metadata”. De twee andere metadatabestanden hebben dezelfde naam als de archiefbestand waar ze bij horen, ook weer met “.metadata” erachter.

In zo’n metadatabestand staat de metadata uit het Toepassingsprofiel Rijk, uitgeschreven in het XML-formaat ToPX:

```
<ToPX>
  <aggregatie>
    <identificatiekenmerk>1015.41</identificatiekenmerk>
    <aggregatieniveau>Dossier</aggregatieniveau>
    <naam>Dossier 1015.41</naam>
    <dekking>
      <inTijd>
        <begin>
          <jaar>1967</jaar>
        </begin>
        <eind>
          <jaar>1981</jaar>
        </eind>
      </inTijd>
    </dekking>
    <openbaarheid>
      <omschrijvingBeperkingen>Openbaar</omschrijvingBeperkingen>
      <datumOfPeriode>
        <datum>2017-01-01</datum>
      </datumOfPeriode>
    </openbaarheid>
  </aggregatie>
</ToPX>
```

Afbeelding: schermafdruck van de XML van een voorbeeld ToPX-metadatabestand

²⁵ Voor de technische specificatie van de exportvoorwaarden, zie <https://www.nationaalarchief.nl/archiveren/kennisbank/voorwaarden-export-naar-e-depot>

In de exportvoorwaarden van het Nederlandse Nationaal Archief staat verder dat de namen van de bestanden en mappen op de harde schijf bepaalde tekens niet mogen bevatten, zoals de dubbele punt (:), de schuine strepen (/ en \) en de asterisk (*). Het geheel moet worden ingepakt tot één bestand en dan naar het e-depot worden verzonden.

Het e-depotsysteem van leverancier Picturae is gebaseerd op het softwarepakket Archivemata. Ook bij dit e-depotsysteem moeten de archieven in een SIP worden aangeleverd, alleen zijn de exportvoorwaarden anders. Een Picturae-SIP kent geen *sidecar*-structuur waarbij elke map en elk bestand een los metadata-bestandje heeft, maar plaatst alle ToPX-metadata in één groot bestand: het Record Information Package-bestand (RIP). Via het element Plaats in de ToPX-metadata weet het systeem wat de hiërarchie van het archiefpakket is.

Opname in het e-depot

Nadat de archiefvormer de archiefbestanden uit zijn bronsysteem heeft geëxporteerd in een formaat dat voldoet aan de exportvoorwaarden van de ontvangende archiefdienst, kan deze SIP naar de archiefdienst worden getransporteerd. Dit kan via internet, via het File Transfer Protocol (FTP), via een eigen netwerk (bijvoorbeeld de “Haagse Ring”, een glasvezelnetwerk tussen ministeries en GemNet, een netwerk tussen gemeenten) of via het versturen van een USB-stick, harde schijf of cd-rom. Welke methode ook gekozen wordt, uiteindelijk moet de SIP in de invoermap van het e-depot terecht komen.

Wanneer de SIP klaar is om opgenomen te worden wordt door een collectiebeheerder van de archiefdienst het ingest-proces gestart. Het e-depot doorloopt het ingest-proces waarbij een aantal inhoudelijke en technische controles wordt uitgevoerd. Wanneer deze succesvol zijn afgerond, dan is de SIP opgenomen in het e-depot. Wanneer er een foutmelding optreedt, dan wordt het opname-proces gestopt en zullen de collectiebeheerder en de archiefvormer samen moeten zoeken naar een oplossing.

Step Progress					
State	Name	Progress	Started	Finished	Messages
	Select		02.05.18 17:47:50	02.05.18 17:49:11	
	Copy SIP		02.05.18 17:49:12	02.05.18 17:49:18	
	Check archiveIdentification exist in XIP Metadata		02.05.18 17:49:19	02.05.18 17:49:21	
	Check Archive Present in Catalogue		02.05.18 17:49:21	02.05.18 17:49:33	
	Identify Metadata Type		02.05.18 17:49:33	02.05.18 17:49:36	
	Transform Metadata		02.05.18 17:49:36	02.05.18 17:49:39	
	Virus Check		02.05.18 17:49:39	02.05.18 17:49:48	
	Metadata Integrity		02.05.18 17:49:48	02.05.18 17:49:51	
	Content Integrity		02.05.18 17:49:51	02.05.18 17:49:54	
	Characterise		02.05.18 17:49:54	02.05.18 17:50:12	View
	Check Collection Addition		02.05.18 17:50:12	02.05.18 17:50:15	
	Fixity Check		02.05.18 17:50:15	02.05.18 17:50:18	
	Transfer Agreement Compliance Check		02.05.18 17:50:18	02.05.18 17:50:21	View
	Encrypted File Check		02.05.18 17:50:21	02.05.18 17:50:24	
	Store Files		02.05.18 17:50:24	02.05.18 17:50:30	
	Store Metadata		02.05.18 17:50:30	02.05.18 17:50:33	
	Store Metadata File		02.05.18 17:50:33	02.05.18 17:50:36	
	Update Search Index		02.05.18 17:50:36	02.05.18 17:50:39	
	Create Memorix File		02.05.18 17:50:39	02.05.18 17:50:51	
	Send Ingest Complete Message to Catalogue		02.05.18 17:50:51	02.05.18 17:50:54	
	Delete SIP		02.05.18 17:50:54	02.05.18 17:50:57	

Afbeelding: voorbeeld van een succesvol doorlopen ingest-proces

Welke stappen precies door het ingest-proces worden uitgevoerd verschilt per softwarepakket dat door het e-depot gebruikt wordt, en per installatie bij de archiefdienst. Een aantal stappen zal bij elke leverancier van e-depots wel worden doorlopen:

- **Virus check**, waarbij gecontroleerd wordt of er virussen en andere *malware* in de bestanden verstopt zit
- **Metadata integrity**, waarbij gecontroleerd wordt of alle metadatabestanden goed zijn ingevuld, en er geen verplichte metadata ontbreekt
- **Characterise**, waarbij van elk bestand in de SIP gekeken wordt welk bestandsformaat het heeft, en vaak zelfs welke versie van het bestandsformaat. Deze informatie wordt door het ingest-proces toegevoegd aan de metadata van de bestanden in de SIP.
- **Fixity check**, waarbij de checksums van de bestanden die in de metadata in de SIP zijn opgenomen worden vergeleken met ter plekke berekende checksums van elk bestand. Wanneer van een bestand de ter plekke berekende checksum niet overeenkomt met de checksum die in de metadata is opgeslagen, dan is er kennelijk bij het exporteren of het transporteren van het archief iets fout gegaan, en moet de SIP opnieuw worden gemaakt en opgestuurd.
- **Store files / store metadata**, waarbij de bestanden daadwerkelijk op het opslagsysteem van het e-depot worden gezet, en waarbij de metadata daadwerkelijk in de database van het e-depot wordt opgenomen.

Nadat het ingest-werkproces is afgerond vindt de laatste visuele controle plaats, waarbij de digitale collectiebeheerder steekproefsgewijs de opgenomen archiefbestanden in het e-depot bekijkt. Daarna wordt de verklaring van overbrenging of uitplaatsing getekend en is de opname voltooid. Hierna kan het verwervingsproces worden afgesloten en kunnen de overgebrachte bestanden in het bronsysteem van de archiefvormer worden verwijderd..

4.3 Oefeningen

Opgave 4.1:

Tijdens de impactanalyse wordt een overbrenging of uitplaatsing uitgevoerd volgens het beschreven proces van beslisdocument maken, metadata-mapping maken, archiefblok exporteren en ingest in het e-depot. Wanneer er enige tijd later een volgende uitplaatsing of overbrenging van dezelfde archiefvormer komt moet dan dit hele proces opnieuw gedaan worden? Zo ja, waarom, zo nee, welke delen zijn te hergebruiken en welke delen moeten wel opnieuw?

Opgave 4.2:

Tijdens de impactanalyse wordt een representatieve dataset bekeken, waarbij een aantal bestanden en de bijhorende mappenstructuur wordt geanalyseerd.

- a) Uit hoeveel mappen en bestanden moet zo'n representatieve dataset bestaan? (tientallen, honderden of duizenden?).

Waar hangt de uiteindelijke keuze tijdens het uitvoeren van een impactanalyse van af?

Opgave 4.3:

Tijdens het uitvoeren van een impactanalyse blijkt dat nergens in het bronsysteem geregistreerd is wie de verantwoordelijke actor is.

- a) Hoe zou je dit tijdens de metadata-mapping oplossen?
- b) Op welk niveau (archief, serie, dossier en/of record) zou je deze informatie vastleggen?

5. Beheer van digitale archieven

In het OAIS-referentiemodel zijn de functies van beheren van metadata en beheren van de bestanden als twee aparte functies opgenomen. In de praktijk zie je dit bij e-depots ook terugkomen in twee aparte processen.

5.1 Beheer van metadata van digitale archieven

Bij het inrichten van de e-depotvoorziening heeft een archiefdienst de keuze waar het de metadata van digitale bestanden gaat beheren: in het e-depot zelf, waar ook de bestanden staan, of in het collectiebeheersysteem waarin ook de metadata van papieren archieven wordt beheerd.

Zowel het Nederlandse Nationaal Archief als het e-depot van Picturae hebben ervoor gekozen om de metadata van opgenomen digitale archieven te beheren in hun centrale collectiebeheersysteem, tezamen met de metadata van papieren archieven.

5.2 Beheer van de digitale bestanden

Het beheer van digitale bestanden omvat twee verschillende processen: het ervoor zorgen dat bestanden niet kwijtraken of stuk gaan, en het zorgen dat bestandsformaten die verouderd raken op tijd worden omgezet naar nieuwe bestandsformaten.

Het ervoor zorgen dat bestanden niet kwijtraken of stuk gaan, wordt gedaan door van elk bestand meerdere kopieën (minstens 3) te bewaren op verschillende fysiek gescheiden locaties. Periodiek wordt van alle bestanden de checksum opnieuw berekend, en vergeleken met de waarde die in de metadata is aangegeven. Wanneer de checksum afwijkt is een bestand kwijt- of stukgeraakt en wordt vanaf één van de backup-locaties een exemplaar teruggezet.

Het omzetten van oude bestandsformaten naar nieuwe bestandsformaten gebeurt via de Preservation-functionaliteit van het e-depot. Wanneer de collectiebeheerder weet welke bestandsformaten verouderd zijn²⁶, in welke bestandsformaten de oude bestanden moeten worden omgezet en met welk softwarepakket de omzetting het beste gedaan kan worden, kan hij het e-depot de opdracht geven deze migratie uit te voeren.

Na de migratie van een oud bestand is het nieuw gemaakte bestand ook opgenomen in het e-depot, voorzien van zijn eigen checksum. De oude versie van het bestand blijft ook bestaan. Mocht er later in de toekomst een betere manier blijken, of wanneer later blijkt dat er tijdens de migratie toch iets is misgegaan, dan is het originele bestand nog altijd aanwezig om een migratie opnieuw uit te voeren.

5.3 Verwijderen van opgenomen archieven

Een speciaal proces dat vooral bij uitgeplaatst archief uitgevoerd zal worden, is het verwijderen van digitale bestanden, dossiers, series of archieven. Bij dit proces moeten zowel de bestanden als (een deel van) de metadata worden verwijderd, en worden dus beide bovenstaande functies aangehaald.

²⁶ Het bijhouden welke bestandsformaten verouderd raken is het onderdeel "Preservation Watch" uit het OAIS-model. Dit kan door de collectiebeheerders zelf worden gedaan, of door iemand met een preservation-rol. Internationaal wordt hieraan gewerkt binnen de [Open Preservation Foundation](#).

Omdat het verwijderen een onomkeerbaar proces is moet het erg zorgvuldig worden uitgevoerd. Wanneer het proces wordt opgestart, bijvoorbeeld omdat de bewaartermijn van een uitgeplaatst dossier is verstreken, dan moet in het werkproces zorgvuldig worden gekeken of er niet alsnog redenen zijn om het dossier langer te bewaren en moet de verwijdering zelf het liefst ‘onder vier ogen’ worden uitgevoerd, waarbij twee personen onafhankelijk van elkaar controleren en goedkeuring geven voor de vernietiging.

5.4 Oefeningen

Opgave 5.1:

Het Nederlandse Nationaal Archief heeft ervoor gekozen om de metadata van de digitale bestanden in het e-depot niet in het e-depot zelf te beheren, maar om ze te beheren in het collectiebeheersysteem waar ook de metadata van de papieren archieven zit.

- a) Noem een voordeel van deze keuze
- Noem een nadeel van deze keuze

Opgave 5.2:

De bestanden in het e-depot worden op tenminste 3 fysiek gescheiden locaties opgeslagen.

- a) Waarom tenminste 3 locaties?
- b) Zou het zin hebben nog meer dan 3 kopieën te bewaren?

Waarom moeten deze locaties fysiek gescheiden zijn?

Wat is volgens jou een veilige minimale afstand tussen deze locaties?

Opgave 5.3:

Bij het verwijderen van bestanden wordt vaak gebruik gemaakt van het ‘vier-ogen’-principe, waarbij een tweede persoon goedkeuring moet geven. Waarom is dat?

Opgave 5.4:

Bij het verwijderen van archiefstukken uit een e-depot kan ervoor gekozen worden om niet alle metadata mee te verwijderen, maar een deel van de metadata te laten staan. Welke reden zou dat kunnen hebben?

6. Beschikbaar stellen van digitale archieven

Zoals in het begin al aangegeven hebben digitale archieven een groot voordeel boven papieren archieven als het gaat om de beschikbaarheid: ze zijn op internet te zetten en daar voor iedereen te raadplegen. Archiefstukken die online beschikbaar zijn gesteld zijn door veel mensen tegelijkertijd te bekijken en nooit “al aan iemand anders uitgeleend”. Born-digital archieven kunnen volledig doorzoekbaar gemaakt zijn, en bij scans kan via OCR doorzoekbare tekst worden toegevoegd. Digitale documenten kunnen ook makkelijk in meer dan één dossier zitten zonder dat ze daadwerkelijk twee keer in het e-depot zijn opgeslagen.

6.1 Beschikbaar stellen via een inventaris

De klassieke inventaris is bij het Nederlandse Nationaal Archief nog altijd de methode om archieven beschikbaar te stellen, ook bij digitale archieven of hybride archieven. Wanneer een dossier één of meer digitale bestanden bevat, dan wordt in de inventaris een linkje opgenomen naar het digitale bestand in het e-depot.

DC/153 GENERIEKE EN GEBIEDSGERICHTE DEELPROGRAMMA'S VAN HET DELTAPROGRAMMA, WAARIN PERIODE 2010-2014 ADVIEZEN EN VOORSTELLEN VOOR DELTABESLISSINGEN EN VOORKEURSSTRATEGIEËN ZIJN VOORBEREID EN UITGEWERKT

DC/159 GENERIEK DEELPROGRAMMA VEILIGHEID. VOORBEREIDING 2010-2014
ADVIEZEN EN VOORSTELLEN DELTABESLISSING WATERVEILIGHEID

DC/256 Synthesedocument Veiligheid. Beschrijving van onderbouwing van de voorgestelde deltabeslissing waterveiligheid tbv Deltaprogramma 2015

DC-2015/1066 Deltaprogramma DP B1 Synthesedocument 2014 Deelprogramma Veiligheid

[DC-2015_1066-1.PDF](#)

DC/365 Brondocumenten en literatuur bij synthesedocument deltabeslissing waterveiligheid

DC-2015/1109 kengetallen kosten-batenanalyse waterveiligheid 21e eeuw

[DC-2015_1109-1.PDF](#)

DC-2015/1969 Tweede Kamer. (2012). Motie van de leden Van Veldhoven en Lucas 17 april 2012 TK nr. 27625 nr. 262. Den Haag.

[DC-2015_1969-1.PDF](#)

Afbeelding: voorbeeld van een deel van de inventaris van het Deltaprogramma-archie²⁷. De namen van de PDF-bestanden zijn links naar de betreffende bestanden in het e-depot

Deze methode van beschikbaar stellen werkt prima voor individuele bestanden, waarbij de gebruiker aan de context en de bestandsnaam kan zien of hij het document wil inzien. Voor een archief dan uit vele honderden of duizenden bestanden bestaat is dit minder geschikt, omdat de gebruiker dan één voor één de bestanden moet aanklikken om te downloaden.

Een methode waarbij een bezoeker in één klik de complete inhoud van een dossier kan downloaden is nog geen onderdeel van de huidige website. In de volgende versie waar nu aan gewerkt wordt gaat dat wel ingebouwd worden.

²⁷ Zie: <http://www.gahetna.nl/collectie/archief/ead/index/eadid/2.16.133>

6.2 Beschikbaar stellen als open data

Een alternatieve methode van beschikbaar stellen is het beschikbaar stellen voor andere computers. Hierbij is niet een mens, maar een machine degene die de informatie ophaalt, eventueel bewerkt, en het dan in één of ander vorm weer aan gebruikers aanbiedt. Deze toegang is niet bedoeld voor menselijk gebruik en ziet er dus ook erg technisch uit. Er wordt dan gebruik gemaakt van technologieën die bekend staan als ‘webservices’ of ‘application programmers interfaces (APIs)’.

Openbare data die hergebruikt mag worden en die geschikt is voor machine-leesbare verwerking heet ook wel ‘open data’.

Wie dapper is kan eens op de volgende link klikken:

http://www.gahetna.nl/archievenoverzicht/oai-pmh?verb=ListRecords&set=naa1&metadataPrefix=oai_ead

Je ziet dan een machine-leesbare lijst met informatie over de archiefinventarissen die het Nederlandse Nationaal Archief aanbiedt als open data. Hierbij wordt gebruik gemaakt van een webservice die voldoet aan de internationale standaard OAI-PMH.

Een andere webservice, nu met de standaard CMIS, is ingebouwd in Preservica, het softwarepakket dat onderdeel is van het Nederlandse e-depot:

<https://anonymous.nlhana:welkominhetedepot@e-depot.nationaalarchief.nl/cmris-server/browser/sdb/root/>

Ook dit is niet bedoeld om leesbaar te zijn voor menselijke gebruikers, maar is door een computer prima te gebruiken.

Het aanbieden van archiefinformatie als open data is in Europa in een richtlijn geregeld²⁸. In de Nederlandse wetgeving is dit uitgewerkt in de Wet hergebruik overheidsinformatie²⁹.

6.3 Beperkingen aan online beschikbaar stellen

Dat iets online beschikbaar gesteld kán worden, betekent niet altijd automatisch dat het ook online beschikbaar gesteld mag worden. Archieven die bij overbrenging een openbaarheidsbeperking hebben gekregen mogen niet zomaar openbaar online gezet worden, en ook de auteursrechtwetgeving en wetgeving over privacybescherming kunnen redenen geven waarom stukken niet online mogen worden gezet. Vaak mogen deze stukken wel offline, in een studiezaal ter beschikking worden gesteld, eventueel na het verlenen van ontheffing van een openbaarheidsbeperking.

6.4 Oefeningen

Opgave 6.1:

Tijdens jouw werkzaamheden maak of beheer je digitale archiefstukken.

- a) Geef een voorbeeld van een soort archiefstuk dat jij maakt of beheert, dat via een website aan het publiek getoond kan en mag worden.
- b) Geef een voorbeeld van een soort archiefstuk dat jij maakt of beheert, dat niet via een website aan het publiek getoond kan of mag worden.

Opgave 6.2:

Kent Curaçao ook wet- of regelgeving die het aanbieden van overheidsinformatie als open-data verplicht of stimuleert? Verwacht je dat er veel vraag is naar het soort informatie dat jij tijdens jouw werk maakt of beheert?

²⁸ Zie: <http://eur-lex.europa.eu/legal-content/NL/TXT/PDF/?uri=OJ:L:2013:175:FULL&from=NL>

²⁹ Zie: <http://wetten.overheid.nl/BWBR0036795/>

7. Nieuwe vormen van informatie

Het voorspellen van de toekomst is altijd moeilijk, en zorgt ervoor dat een tekst snel verouderd kan raken. Mocht je dit later herlezen, hou er dan rekening mee dat dit een tekst is uit 2018. En omdat het vooral vragen zijn waarop nu nog geen eenduidig antwoord op gegeven kan worden, is dit ook geen examenstof.

De informatie die we in de praktijk nu tegenkomen bij het aansluiten van een archiefvormer op een e-depot, is al een flink aantal jaar geleden gevormd. Het is dan ook vaak een digitale manier van werken die erg lijkt op de oude papieren manier van werken: het gaat om digitale documenten met bijlagen, ze zitten samen met andere documenten in dossiers, de dossiers zitten in series of rubrieken, en die vormen samen een archiefblok.

Maar doordat we zelf als archiefmedewerkers ook archieven vormen, weten we dat de informatie-objekten die vandaag de dag gecreëerd wordt lang niet altijd meer een document-vorm hebben.

Databases

In het allereerste hoofdstuk over 'Wat is archief' werd al genoemd dat databases met bijvoorbeeld meetgegevens gebruikt kunnen zijn in werkprocessen, en daarmee archiefstukken zijn. Maar een database heeft niet de vorm van een document. Ook kunnen specifieke regels in een database informatie bevatten die bewaard moet worden, terwijl andere informatie in die database vernietigd moet worden. Veel databases worden gebruikt in combinatie met een sjabloon, waarbij labels in het sjabloon duiding geeft aan de informatie in de database. En wanneer is informatie in een database 'afgesloten' zodat het na 20 jaar overgebracht kan worden?

Websites

In de archiefwetgeving staat dat alle in een werkproces gecreëerde informatie archief is. Websites vallen daar ook onder, meestal in het werkproces van de afdeling die zich bezighoudt met publiciteit en voorlichting. Websites hebben de eigenschap dat ze heel makkelijk en snel aangepast kunnen worden en er dus wel dagelijks of meerdere keren per dag een aangepaste inhoud is. Ook zijn er gepersonaliseerde websites waarbij de getoonde inhoud afhankelijk is van wie is ingelogd.

Naast de veranderende inhoud van een website, geeft ook de vorm zijn eigen uitdagingen. Een webpagina is niet één document, maar samengesteld uit losse onderdelen: teksten, afbeeldingen, vormgeving, bijlagen, audio- en videobestanden, en links naar andere webpagina's binnen en buiten de website en interactieve elementen waarbij de inhoud afhangt van wat één of meer bezoekers van de website doen. Bij het archiveren van een website is het van belang al deze componenten zo goed mogelijk op te slaan, te beheren en weer te kunnen weergeven zodat aan alle vier doelen van archivering (eigen bedrijfsvoering, verantwoording, recht- en bewijsvinding en cultuurhistorisch) kunnen worden behaald.

E-mail

Veel communicatie die relevant is in besluitvorming vindt plaats via e-mail. Nu is e-mail niet echt nieuw, het bestond 20 jaar geleden ook al. Wel is het veel minder lang gebruikelijk dat e-mail informatie bevat die niet geprint in een dossier wordt opgeslagen. In theorie zou e-mail geen grote uitdaging hoeven bieden: e-mails lijken erg op documenten, al dan niet met bijlagen en gerelateerd aan andere e-mails. Maar in de praktijk komt het heel veel voor dat de informatie blijft hangen in het "Postvak In" en de "Verstuurde items". De e-mails worden dan niet in het DMS opgeslagen bij het dossier waarin het thuis zou horen. In mailboxen zit vaak zakelijke en privémails door elkaar heen, en

ze zitten vol met privacygevoelige persoonsinformatie. Wat te doen met al die mailboxen van medewerkers die uit dienst gaan, en hoe hou je ze toegankelijk wanneer er een verzoek binnenkomt op basis van openbaarheid van bestuur?³⁰ Binnen Nederland lopen nu pilots om, naast het opslaan van e-mails in de DMS'en bij de dossiers waarin ze thuishoren, van sleutelpersonen ook de hele mailbox te bewaren. Hoe goed de informatie in die mailboxen weer te raadplegen is, en of daarmee het overheidshandelen voldoende kan worden verantwoord, moet uit de pilots blijken.

Messaging apps en sociale media

Naast e-mail wordt tegenwoordig heel veel informatie uitgewisseld via messaging apps (bijv. WhatsApp, Signal, Telegram) en via sociale media (bijv. Facebook, Twitter, Instagram, Snapchat). In de definitie van archief en archiefbescheiden staat heel duidelijk dat alles wat een overheidsorgaan als procesgebonden informatie ontvangt of maakt en verstuurt, archief is. Ongeacht de vorm. Contact met burgers of met andere overheidsmedewerkers kan daar zeker onder vallen.

Deze vormen van informatie hebben een eigen uitdaging: de informatie staat niet in een informatiesysteem waarvan de overheid het eigendom heeft. De informatie staat op de servers van Facebook, WhatsApp en Twitter zelf. Dat die organisaties over 20 jaar nog bestaan, en dat de informatie er in de juiste vorm uitgehaald kan worden, is nog maar de vraag. En hoe garanderen we de authenticiteit? Hoe weten we zeker dat de informatie die gearchiveerd wordt ook inderdaad de informatie is die destijds gepubliceerd was, wanneer we geen inzicht hebben in de kwaliteitseisen die commerciële partijen aan hun informatiesystemen stellen.

Verantwoordelijkheid over algoritmes

Het begrip 'verantwoordelijkheid over algoritmes'³¹, beter bekend onder de Engelse term 'algorithmic accountability', houdt in dat bedrijven en overheden die algoritmes inzetten bij het maken van beslissingen, moeten kunnen verantwoorden hoe die beslissing tot stand is gekomen. Zelflerende algoritmes worden tegenwoordig ingezet om te voorspellen in welke wijk verhoogde kans is op criminaliteit, bij wie er een verhoogde kans is op fraude, of welke auto op de snelweg staande gehouden moet worden omdat er iets mis zou kunnen zijn.

Voor een overheid of een bedrijf zijn deze algoritmes heel interessant, en kan ze veel geld schelen. Maar hoe kun je je als burger of klant verdedigen tegen de beslissing van een algoritme? Wanneer bij een persoon om de week een keer een inval wordt gedaan omdat een algoritme blijft aangeven dat er een grote kans is dat hij fraudeert, kan dan nog achterhaald worden op basis van welke gegevens en met welke rekenregels het algoritme tot die conclusie komt?

Blockchain

Dé hype van 2017 was toch wel blockchain, en de bijhorende digitale munt Bitcoin. Een blockchain is een soort database die niet op één computer staat, maar die bij alle deelnemers gelijktijdig staat, en min of meer gelijktijdig dezelfde informatie bevat. Ook geeft het de garantie dat de informatie die in een blockchain is opgeslagen niet meer kan worden gewijzigd. Ook hier spelen overheden graag op in en zijn in Nederland volop pilots aan het doen.

Voor de archiefwereld is hierbij een dubbele vraag: zit er in blockchains informatie die gearchiveerd moet worden omdat het procesgebonden informatie is die niet ergens anders te vinden is? En in hoeverre biedt Blockchain-technologie kansen voor de archieven om zelf informatie in op te slaan?

³⁰ Voor Nederland: de [Wet openbaarheid van bestuur](#), voor Curaçao is er de [Landsverordening openbaarheid van bestuur](#).

³¹ Een algoritme is een set rekenregels waarmee een computer een berekening maakt.