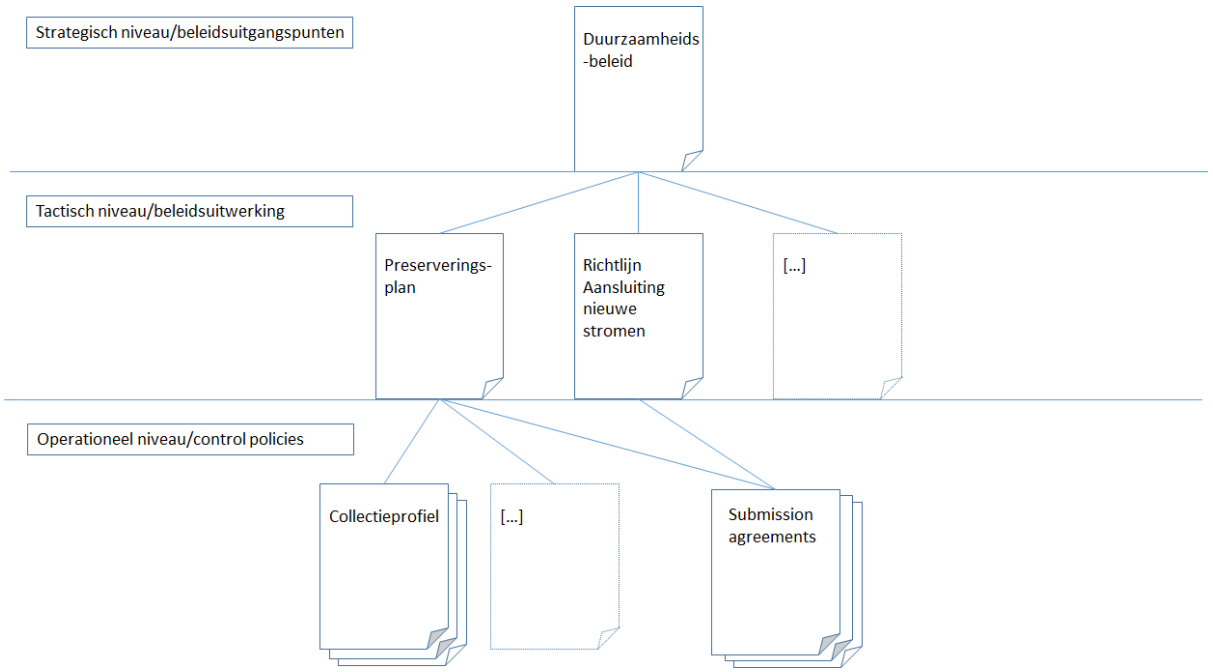
Preserveringsplan 2019-2022

Inleiding

Het doel van dit document is om nadere invulling te geven aan het vastgestelde duurzaamheidsbeleid van de KB voor de periode 2019-2022[1]. Waar het duurzaamheidsbeleid de uitgangspunten op strategisch niveau heeft vastgelegd geeft dit document richting aan de uitwerking van genoemde uitgangspunten op een tactisch niveau. Het niveau van het huidige document correspondeert tevens met wat binnen de NCDD-catalogus genoemd wordt als het niveau van de beleidsuitwerking. Het duurzaamheidsbeleid formuleert in die terminologie dan de beleidsuitgangspunten. Het huidige document heeft niet precies dezelfde opbouw als het duurzaamheidsbeleid omdat we verwachten dat het een andere doelgroep zal aanspreken en we het belangrijk vinden dat het gelezen kan worden als zelfstandig document.

De onderwerpen die binnen dit document uitgewerkt worden betreffen die randvoorwaardelijke zaken uit het duurzaamheidsbeleid die betrekking hebben op het duurzaam bewaren van de collectie. Om dit mogelijk te maken wordt er ook verwezen naar een aantal bredere randvoorwaarden zonder welke de garantie van duurzaamheid niet mogelijk is zoals de juridische en organisatorische kaders[2]. Deze zijn ontleend aan het duurzaamheidsbeleid maar zijn breder van scope dan uitsluitend duurzaamheid en zullen daarom niet uitgewerkt worden in het preserveringsplan. Het huidige plan beoogt niet alle informatie over duurzaamheid op tactisch niveau te bundelen maar bevat wel verwijzingen naar andere documenten op hetzelfde niveau waar zaken genoemd staan die hier niet of niet geheel uitgewerkt zijn. Tevens bevat dit document ook verwijzingen naar documenten op operationeel niveau of suggesties voor het opstellen van documenten op dit niveau als praktische invulling van punten uit de beleidsuitwerking. Zie ter verduidelijking onderstaande illustratie waarin de voorgestelde hiërarchie van documenten getoond wordt. Het duurzaamheidsbeleid vormt de basis van alle stukken op het tactische niveau. Deze vormen op hun beurt weer de basis voor stukken op het operationeel niveau. Het kan zijn dat een operationeel stuk zoals de submission agreements invulling geeft aan uitgangspunten uit meerdere stukken op tactisch niveau.

[1] [Definitieve locatie volgt nog]
[2] Ibid, zie bijvoorbeeld hoofdstuk 2.6 Rechten en 2.9 Organisatie



Het is ten slotte goed om te beseffen dat dit document niet een concreet preserveringsactieplan bevat. De uitgangspunten voor preserveringsplanning als proces worden wel beschreven in het tactische beleid maar concrete actieplannen moeten opgesteld worden op operationeel niveau, gebaseerd op risico-analyse en preservation watch. Vanuit het OAIS-model gezien is dit document het beleid vanuit Administration op basis waarvan de functionele entiteit Preservation Planning kan werken aan voorstellen voor actieplannen.

Verhouding van dit document tot andere beleidsdocumenten en ISO-16363

Dit document is gerelateerd aan een aantal beleidsdocumenten zoals deze zijn vastgesteld binnen de KB.

Ten eerste het duurzaamheidsbeleid. Dit document stelt de strategische uitgangspunten waar het huidige document een tactische invulling van is. Het strategische beleid kan daarmee beschouwd worden als het plan dat in ISO-16363 ‘Preservation Strategic Plan’ wordt genoemd en is daarmee een invulling van eis 3.1.2.[3] Het huidige document is op het tactisch niveau dan een eerste invulling van eis 3.3.2 waarin gesteld wordt dat een repository beleidsdocumenten moet hebben die een uitwerking zijn van het strategische plan.[4] In ISO-16363 worden dergelijke documenten ‘preservation policies’ genoemd maar om verwarring te voorkomen met het strategische beleid zullen we in dit stuk naar dergelijke documenten verwijzen met de benaming ‘preserveringsplan’.

[3] "September 2011 Magenta Draft of ISO 16363", https://public.ccsds.org/pubs/652x0m1.pdf: blz. 3-1
[4] Ibid. blz. 3-6

Als tweede kan genoemd worden de Content strategie van de KB. Dit stuk bevat in hoofdlijnen de doelen en uitgangspunten over collectievorming en toegang en is daarmee randvoorwaardelijk voor het maken van keuzes over wat er duurzaam bewaard dient te worden. Dit stuk vervult daarmee eis 3.1.3 waarin gesteld wordt dat er een collectiebeleidstuk beschikbaar moet zijn.

Ten slotte kan er verwezen worden naar verschillende stukken op het gebied van security, privacy en risico-management zoals het Bedrijfshulpverleningsplan[5], Bedrijfsnoodplan[6], Informatiehulpverleningsplan[7] en het Collectiehulpverleningsplan[8]. Deze documenten zijn grofweg een invulling van eis 5.1.2 waarbinnen gesteld wordt dat er voorzieningen moeten zijn om veiligheidsrisico's op het gebied van systemen, personeel en fysieke condities onder controle te kunnen houden.[9] Zie voor meer uitwerking het hoofdstuk Informatieveiligheid hieronder.

Bovengenoemde documenten vormen het bredere kader waarbinnen er nagedacht kan worden over het duurzaam bewaren van de collectie zelf. Om te zorgen dat processen om de data authentiek en toegankelijk te houden ook echt effectief zijn, is het van belang dat de omgeving waarbinnen deze processen functioneren ook op orde is.
Dit betekent dat er rekening gehouden moet worden met personele, financiële en infrastructurele aspecten die indirect impact hebben op de duurzaamheid van de collectie. In hoofdlijnen zijn deze zaken verwerkt in bovengenoemde documenten maar voor de toekomst is het belangrijk dat er ook hier het aspect van digitale duurzaamheid in opgenomen wordt. Zo is het Duurzaamheidsbeleid een goede invulling van eis 3.1.2 zoals hierboven genoemd maar moeten er nog documenten opgeleverd worden om ook te kunnen voldoen aan de onderliggende eisen die stellen dat er bijvoorbeeld een successieplan is en dat er maatregelen zijn getroffen voor disaster recovery.[10]

Ten slotte is het op het gebied van infrastructuur ook van belang dat er een duidelijk gedocumenteerd proces is waarbij veranderingen aan de infrastructuur in gang gezet kunnen worden op basis van voorziene risico's op het gebied van duurzaamheid van de collectie. Daarbij is het ook van belang dat er een financiële buffer is om noodzakelijke vernieuwingen te kunnen bekostigen.[11]

Beleidsuitwerking van duurzaamheidsthema's

Hieronder volgt de tactische invulling van de duurzaamheidsthema's uit het duurzaamheidsbeleid. De invulling van de begrippen moet regelmatig getoetst worden en de gekozen oplossingen opnieuw geëvalueerd vanwege veranderingen op het gebied van

[5] https://plein.kb.nl/documents/51711
[6] https://plein.kb.nl/documents/51710
[7] [locatie volgt nog]
[8] https://plein.kb.nl/documents/51712
[9] "September 2011 Magenta Draft of ISO 16363", https://public.ccsds.org/pubs/652x0m1.pdf: blz. 5-12
[10] Ibid. 3.1.2.1, blz. 3-2
[11] Ibid. Zie eisen onder 5.1.1, Blz. 5-1 ff

wetgeving, software-ontwikkeling, software-contracten, IT-infrastructuur, verwachtingen van gebruikers en wensen van depotgevers.

Integriteit
Onder het begrip integriteit verstaan we de garantie dat objecten en collecties volledig zijn en dat wijzigingen gecontroleerd en gedocumenteerd verlopen.

Er zijn meerdere niveau's van integriteit die we controleren op basis van verschillende maatregelen.

-Bit-integriteit[12]
-Versie-integriteit[13]
-IP-integriteit[14]
-Informatie-integriteit[15]
-Collectie-integriteit[16]

Bit-integriteit
Met het controleren van de *bit-integriteit* van bestanden willen we aantonen dat een digitaal bestand zoals deze is opgeslagen binnen de IT-infrastructuur identiek is aan de aangeleverde publicatie of de laatste versie. De bit-integriteit van bestanden wordt gecontroleerd aan de hand van een checksum. Deze checksum wordt meegeleverd door depotgevers bij de aanlevering. Wanneer het zo is dat er geen checksum meegeleverd kan worden moet er tenminste vooraf een ander vergelijkbaar mechanisme zijn afgesproken met de depotgever om een betrouwbare overdracht te garanderen. De checksum wordt dan in ieder geval alsnog berekend en opgeslagen na ontvangst. Bij wijzigingen aan een bestand is het van belang dat er ook altijd een nieuwe checksum wordt berekend en opgeslagen.
De checksum wordt per bestand bijgehouden. De IT-infrastructuur is zo ingericht dat bestanden regulier gecontroleerd worden en op basis van checksum geverifieerd worden om bit-rot te detecteren. Rapportages over dit proces worden ter review op reguliere basis geleverd. Wanneer er fouten worden gedetecteerd binnen het systeem wordt door middel van self-healing en replicatie gezorgd dat bestanden integer blijven en hiervan worden ook rapportages gemaakt.
Bit-integriteit betekent ook dat bits opgeslagen worden zoals ze ontvangen zijn en dat ze niet ongecontroleerd gewijzigd zijn. Daarom is het ook belangrijk dat de data niet geëncrypteerd, gecomprimeerd of gededupliceerd wordt op opslag niveau, zodat de bit-integriteit bewaard blijft.

Versie-integriteit
Vanuit het beleidsuitgangspunt om altijd het origineel te bewaren en alle versies van een publicatie is het van belang om ook de *versie-integriteit* te bewaken. Bestanden kunnen namelijk origineel zijn zoals oorspronkelijk ontvangen van de depotgever, een nieuwe versie

[12] Dit is een invulling van ISO-16363 eis 4.4.1.2 en 5.1.1.3
[13] Dit is een invulling van ISO-16363 eis 4.4.1.1 en vergelijkbaar met de daar als voorbeeld gestelde methode van implementatie. Zie ook eis 5.1.2.
[14] Dit is een invulling van ISO-16363 eis 4.1.5 (SIP) en 4.2.1 en 4.2.9 (AIP)
[15] Dit is een invulling van ISO-16363 eis 4.1.2 (SIP) en 4.2.5 (AIP)
[16] Dit is een invulling van ISO-16363 eis 4.2.9

geleverd door de depotgever om een fout te herstellen. In dergelijke gevallen is het belangrijk voor de integriteit van het materiaal dat relaties tussen versies van een intellectuele entiteit bewaard blijven. Door relaties tussen versies te bewaren en inzichtelijk te maken is het mogelijk om de chain of custody van een object te beheren. Hierdoor kan aan depotgevers en de designated community aangetoond worden dat versies van een object integer zijn en terug te herleiden tot het origineel.

IP-integriteit
Een ander niveau van integriteit is de *IP-integriteit*. Hierbij gaat het er om dat er in de verschillende stadia die Information Packages doorlopen wordt gecontroleerd op volledigheid. Bij ingest wordt gecontroleerd dat alle verwachte bestanden in de SIP ook daadwerkelijk geleverd zijn, gebaseerd op afspraken met depotgevers. Voor de AIP geldt dat alle bestanden zoals deze oorspronkelijk zijn meegeleverd terug te herleiden zijn binnen een of meerdere AIPs. Hetzelfde geldt voor de DIPs. Het is hierbij belangrijk op te merken dat er vanuit het perspectief van duurzaamheid gezien geen onderscheid wordt gemaakt tussen hoofdbestanden en aanvullende bestanden. Wat er tot de IP behoort wordt vastgelegd in een manifest-bestand dat ook dient als referentie bij volledigheidscontroles bijvoorbeeld tijdens ingest of bij migraties.

Informatie-integriteit
Daarnaast is er ook het niveau van de *informatie-integriteit* die bewaakt moet worden. Bij dit begrip gaat het er om zeker te stellen dat alle representatie informatie die nodig is om een object te interpreteren ook duurzaam bewaard blijft en persistent gerelateerd blijft aan het object. Dit kan door het opslaan van extra informatie als metadata of als document. In de praktijk wordt ook documentatie zoals benodigd om de informatie binnen de AIP te begrijpen opgeslagen met de AIP of als afzonderlijke AIP met verwijzing.

Bovengenoemde zaken zoals de samenstelling van een AIP of de manier waarop er versies van een object bestaan verschillen in de praktijk per collectie. Als hulpmiddel om deze verschillen inzichtelijk te maken en vast te leggen zullen collectieprofielen en IP-schema's gemaakt worden. Dit is van belang om zeker te zijn dat er één waarheid is omtrent wat behoort tot de AIP en wat de status is van objecten binnen een collectie.
Laatstgenoemde punt betreft het overkoepelende niveau van de *collectie-integriteit*, waarbij zeker gesteld moet worden dat de collectie als geheel ook integer is. Het gaat hierbij om de controle dat alle objecten die binnen een collectie aanwezig zouden moeten zijn ook een definitieve status hebben gekregen. Een definitieve status kan bereikt worden doordat is vastgesteld aan het einde van het ingest-proces dat een volledige gecontroleerde AIP is opgeslagen op de definitieve locatie. Of dat er op basis van gedocumenteerde procedures voor is gekozen een object definitief niet op te nemen waarbij er in bepaalde gevallen ook een tombstone-record wordt bewaard om deze keuze voor de lange termijn inzichtelijk te houden.

Authenticiteit

Waar de hierboven beschreven mechanismes rondom het begrip integriteit de vraag beantwoorden: "is het materiaal volledig en niet onbedoeld gewijzigd?", gaat het er bij het

begrip authenticiteit om dat aangetoond kan worden dat een object is wat het lijkt te zijn[17] of zou moeten zijn volgens Submission Agreements en dat stappen binnen de IP-lifecycle volgens vaste procedures zijn verlopen. Hierbij moeten verschillende attributen van een object belicht worden namelijk: intentie, herkomst en geschiedenis. Deze aspecten hebben ook een sterke afhankelijkheid met het ingestproces omdat hier al rekening gehouden moet worden met het verkrijgen van alle informatie die nodig is om alle aspecten van authenticiteit te kunnen garanderen.

Intentie
Het is belangrijk dat aangetoond kan worden dat het object overeenkomt met hetgeen de producer bedoeld had om aan te leveren.[18] Bovengenoemd concept van integriteit speelt hierbij ook een rol doordat onvolledigheid van een object kan resulteren in een situatie waarbij het object niet meer weergegeven kan worden zoals bedoeld. Tegelijkertijd gaat het hier ook om de juistheid van de bestanden inhoudelijk gezien. Het betekent niet per se dat er geen fouten mogen voorkomen in een bestand of in de naamgeving maar dat tenminste de data zoals deze opgeslagen moet worden ook daadwerkelijk geleverd is. Wanneer een bestand bijvoorbeeld een .doc-extensie heeft maar eigenlijk een PDF is hoeft dit niet te betekenen dat een bestand niet authentiek is. Het kan dan een preserveringsactie zijn om het bestand te hernoemen. Als echter een bestand een foutmelding bevat in plaats van inhoudelijke metadata dan komt het object niet overeen met de intentie van de producer. Deze laatste categorie is relevant om te controleren in het kader van authenticiteit. Om dit te bereiken wordt in de huidige ingest-processen gebruik gemaakt van software om bestanden te identificeren. Het vergroten van de kennis rondom formaten is belangrijk om toekomstige risico's op het gebied van toegankelijkheid van het formaat te kunnen signaleren, zoals hieronder verder beschreven. Dit aspect van authenticiteit is ook belangrijk in het geval van formaatmigratie waarbij aangetoond moet worden dat een nieuwe vorm van het object nog zoveel mogelijk overeenkomt met de vorm van het object zoals dit door de maker bedoeld was op basis van significante karakteristieken.

Herkomst
Authenticiteit kan ook aangetoond worden door de *herkomst* van het object op te slaan in metadata binnen de AIP. Wat er tot de herkomst van een object behoort wordt bepaald in overleg met de producer die zelf ook een onderdeel uitmaakt van de herkomst. Het verifiëren van de producer gebeurt momenteel binnen de ingest-processen zodat altijd duidelijk is van welke partij een object afkomstig is.[19] Tevens garanderen we dit aspect van authenticiteit door het afsluiten van contracten met depotgevers en slaan we de informatie hiervan en verwijzingen naar de aanleverende partij op binnen het object.

Geschiedenis
Ten slotte moet ook het vastleggen van de *geschiedenis* van het object bijdragen aan de authenticiteit van het object. Deze vastlegging wordt gedaan door het toevoegen van metadata rondom specifieke events die hebben plaatsgevonden tijdens de hele levensloop

[17] Zo kan een bestand een bepaald formaat lijken afgaande op file-extensie terwijl de data in het bestand in een formaat is dat niet overeenkomt met de extensie, bijvoorbeeld omdat het verkeerd hernoemd is.
[18] Dit is een invulling van ISO-16363 eis 4.1.3, 4.1.5
[19] Dit is een invulling van ISO-16363 eis 4.1.4

van het object bij de KB. De events worden momenteel vastgelegd in het manifest-bestand dat bij elke AIP aanwezig is.[20]

Minimale ingest

Momenteel worden veel controles gedaan tijdens het ingestproces waarbij geconstateerde fouten soms ook leiden tot het niet opslaan van materiaal. In de toekomst is het de bedoeling om dergelijke processen zoveel mogelijk na ingest in te plannen en uit te breiden zodat er bijvoorbeeld meer technische metadata beschikbaar is voor alle formaten en dat deze ook gekoppeld zijn aan een extern register.[21] Daarnaast heeft dit ook als voordeel dat materiaal direct veilig wordt opgeslagen en niet blijft hangen in fout-mappen op een verwerkingsserver.

Dit is wat we hier minimale ingest noemen. Het doen van controles willen we handhaven en er blijven controles uitgevoerd, maar de meeste controles verschuiven naar de fase binnen de IP-levenscyclus van SIP naar AIP om op deze manier het veiligstellen van het materiaal te prioriteren.[22] Het doen van de controles gebeurt dan nadat het materiaal veilig is opgeslagen en eventuele foutafhandeling kan uitgevoerd worden als een gedocumenteerde wijziging waarbij het oorspronkelijke bestand bewaard blijft. Ook kunnen de events hiervan dan vastgelegd worden, zodat de authenticiteit gegarandeerd kan worden. Dit is één van de redenen waarom objecten zo snel mogelijk geïngest moeten worden. De events moeten ook geraadpleegd kunnen worden, zodat de chain of custody geëvalueerd kan worden. Het moet duidelijk zijn wat er met een object is gebeurd en daarom moeten alle bestanden bewaard blijven. Daarnaast wordt er een audit log bijgehouden om alle acties ook te documenteren. Ook als objecten verwijderd worden, moet er metadata bewaard blijven zodat duidelijk is wat er ooit was.

De controles tijdens ingest moeten er op gericht zijn om het beheer over te nemen van een object, dat wil zeggen deze moeten voldoende zijn om bit-preservation te kunnen garanderen. Dit betekent dan ook dat het object in eerste instantie is opgeslagen op kennisniveau 1 (zie hieronder). De benodigde identificatie en validatie van de bestandsformaten binnen de SIP zoals deze nodig is voor functionele preservering wordt na ingest gedaan op de AIP waarna kennisniveau en preserveringsniveau aangepast kunnen worden. De bedoeling is om het toekomstige ingestproces te optimaliseren zodat deze scheiding tussen acties voor bit-preservering en functionele preservering duidelijker wordt. Daarnaast is het ook van belang om in de toekomst op een vaste manier te documenteren hoe SIPs worden omgezet naar AIPs en om dit als een procedure vast te leggen voor raadpleging.[23] Uiteindelijk moet dit proces ervoor zorgen dat bepaalde vereisten vanuit het ingest-proces zoals dat materiaal niet versleuteld, gecomprimeerd of dubbel mag zijn ondervangen worden door middel van geautomatiseerde controles.

Duurzame toegankelijkheid

Er zijn verschillende aspecten van toegankelijkheid die gegarandeerd moeten zijn om deze te kunnen beschouwen als duurzaam. De gebruiker moet materiaal kunnen vinden op basis

[20] Dit is een invulling van ISO-16363 eis 4.1.8 (SIP) en 4.2.10 (AIP)
[21] Dit zou bijdragen aan het aantonen van compliance met eisen zoals ISO-16363 eis 4.2.5
[22] Ons begrip van minimale ingest is ontleend aan de invulling van de Deense Staatsbibliotheek: https://en.statsbiblioteket.dk/about-the-library/projects-1/MinEffortIngest_iPRES2015.pdf waar ook wordt aangetoond dat een dergelijk proces in overeenstemming is met de uitgangspunten van OAIS
[23] Dit is een invulling van ISO-16363 eis 4.2.2

van metadata[24], een gevonden object moet weergegeven kunnen worden[25], de inhoud van het object moet begrijpelijk zijn[26], de gebruiker moet kunnen bepalen in hoeverre een object integer en authentiek is en het object moet uitgeleverd kunnen worden als een DIP[27].
Dit komt overeen met de begrippen vindbaar, leesbaar, interpreteerbaar, betrouwbaar en beschikbaar.[28] De definitie van het begrip 'betrouwbaar' is dus ook direct verbonden met de invulling van de begrippen integriteit en authenticiteit zoals hierboven beschreven. Integriteit en authenticiteit zijn daarmee randvoorwaardelijk voor duurzame toegankelijkheid.

Al deze punten bevatten ook een juridische afhankelijkheid omdat voor al deze processen gewaarborgd moet zijn dat er voldoende rechten zijn om acties uit te voeren, dat deze rechten inzichtelijk zijn en dat deze ook gehandhaafd worden binnen de processen.

Momenteel wordt dit verwezenlijkt door afspraken hierover vast te leggen in contracten met depotgevers en tevens door rechtenmetadata vast te leggen en dit, waar relevant, te vertalen naar toegangsrechten binnen de systemen. Het proces rond contractbeheer is nu een losstaand proces, dit moet in de toekomst een beter geïntegreerd proces zijn, zodat duidelijker is welk contract met welke publicatie verbonden is en zo ook de rechten die ontleend kunnen worden aan het contract voor duurzaamheid duidelijk zijn vastgelegd en ook de onderliggende afspraken duurzaam bewaard blijven.[29] De processen rond duurzaamheid moeten verder geformaliseerd en gedocumenteerd worden, waarbij er dan ook uitgewerkt wordt hoe deze processen kunnen worden aangepast in reactie op veranderingen. Voor de invulling hiervan zal bekeken worden in hoeverre ISO 9001 een handvat biedt.

Of objecten beschouwd kunnen worden als toegankelijk, wordt bepaald aan de hand van input uit drie informatiebronnen:

1. Afspraken met depotgevers (submission agreements)[30]
2. Gebruikersonderzoek (monitoring designated communities)[31]
3. Preservation watch (risico-analyse onder andere op basis van technology watch en monitoring designated communities)

De input vanuit gebruikersonderzoek en preservation watch zal gebruikt worden om te toetsen in hoeverre de aspecten van duurzame toegankelijkheid zoals hierboven beschreven door de tijd heen nog waargemaakt kunnen worden. De afspraken met depotgevers en informatie uit preservation watch vormen de input voor de juridische randvoorwaarden die ook meegewogen moeten worden bij toekomstige oplossingen om toegankelijkheid te waarborgen.

---

[24] Dit is een invulling van ISO-16363 eis 4.5.1
[25] Dit is een invulling van ISO-16363 eis 4.3.2/4.2.7
[26] Dit is een invulling van ISO-16363 eis 4.3.2/4.2.7
[27] Dit is een invulling van ISO-16363 eis 4.6.2
[28] Zoals geformuleerd binnen de definitie van duurzaam toegankelijk beschreven in Duto: https://www.nationaalarchief.nl/archiveren/kennisbank/duurzaam-toegankelijk
[29] Dit is een invulling van ISO-16363 eis 3.5.1
[30] Dit met name voor het juridische aspect van toegankelijkheid
[31] Het testen van begrijpelijkheid wordt bijvoorbeeld genoemd in ISO-16363 eis 4.2.7

De afspraken met depotgevers gaan in de toekomst verkend en vastgelegd worden volgens de stappen van het PAIMAS-model. Bij preservation watch is risico-management een belangrijk onderdeel om de duurzaamheid van de digitale collectie veilig te stellen. Naast de duurzaamheid van de collectie op zich is risico management in de KB brede zin belangrijk voor de duurzaamheid van de collectie Daarnaast zal er ook technology watch systematisch ingericht moeten worden. Hier zal monitoring komen die toegespitst is op het tijdig signaleren van relevante technologie- en software-ontwikkelingen. Voor gebruikersonderzoek zal aansluiting gezocht worden bij bestaande initiatieven binnen de KB op dit gebied zodat deze uitgebreid kunnen worden op een manier die relevant is voor duurzaamheid. Ook zal de formulering van de designated community uitgebreid worden. Tenslotte zullen ook interne processen voor bestandsanalyse van formaten een belangrijke bron worden voor preservation watch. Zie tevens het hoofdstuk Kennisniveaus dat verder beschrijft hoe hier vorm aan gegeven gaat worden.

Preserveringsstrategieën
Binnen het duurzaamheidsbeleid worden twee preserveringsniveaus onderscheiden: bit-preservering en functionele preservering. Het doel is om uiteindelijk over alle bestanden functionele preserverering toe te passen. De preserveringsniveaus zoals deze toegekend kunnen worden aan objecten zijn dus een momentopname.

Bij *bit-preservering* gaat het erom dat de bits binnen een bestand aantoonbaar onveranderd zijn sinds het moment van vaststelling. Bovengenoemde maatregelen om de integriteit, authenticiteit en toegankelijkheid van objecten te waarborgen zijn hierbij belangrijk. Dat iets opgeslagen is op het niveau van bit-preservation wil dus zeggen dat er nog steeds zeer hoge eisen worden gesteld op het gebied van duurzaamheid maar door de aard ervan kunnen niet alle aspecten van integriteit, authenticiteit en toegankelijkheid volledig waargemaakt worden. Specifiek zal voor integriteit de informatie-integriteit niet volledig gegarandeerd kunnen worden. Wat betreft toegankelijkheid zullen de aspecten leesbaar en interpreteerbaar niet gegarandeerd kunnen worden. Dit niveau is dus duurzaam voor de korte termijn omdat de levensduur van vorm en inhoud rechtstreeks gekoppeld is aan de levensduur van het formaat en de software-omgeving waarbinnen het formaat functioneert.

Bij *functionele preservering* gaat het erom dat een representatie van het oorspronkelijke bestand toegankelijk gemaakt kan worden. Ook hier spelen de definities van integriteit en authenticiteit een belangrijke rol maar wordt er daarnaast extra aandacht besteedt aan het bewaren van vorm en inhoud los van het formaat. Bit-preservering maakt altijd een onderdeel uit van functionele preservering maar daarnaast worden er bij functionele preservering extra maatregelen getroffen voor instandhouding van de toegankelijkheid door de tijd heen. Dit niveau is daarom duurzaam voor de lange termijn.
Binnen het beleid is als doel gesteld om de focus te verleggen van uitsluitend bit-preservering naar functionele preservering. De eerste stap in dit proces is het vastleggen van *preserveringsstrategieën*[32], *preserveringsniveaus*, het uitwerken van een proces voor *preserveringsplanning* en inrichting van de vereiste monitoring mechanismes zoals hierboven benoemd onder toegankelijkheid.

[32] Dit is een invulling op hoofdlijnen van ISO-16363 eis 4.3.1

De twee belangrijkste strategieën voor functionele preservering zijn emulatie en formaatmigratie. Vanwege de diverse collectie van de KB zullen beide relevant zijn om te onderzoeken in het kader van duurzame toegankelijkheid.

Bij *emulatie* wordt de toegankelijkheid gegarandeerd door in stand houding van een software-omgeving waarbinnen een oorspronkelijk bestand toegankelijk blijft. Hierbij kan het ook van belang zijn om de oorspronkelijke hardware en/of software voor een vastgestelde tijd toegankelijk te houden.

Bij *formaatmigratie* wordt de inhoud van een object overgezet naar een nieuw formaat zodat de inhoud toegankelijk blijft. Voor een goed begrip van deze term is het belangrijk om in navolging van OAIS onderscheid te maken tussen de vier verschillende types van migratie. Deze zijn te onderscheiden op basis van het doel dat ze beogen en de impact die ze hebben op de data die gemigreerd wordt.

1. Refreshment, het migreren van objecten naar eenzelfde medium zonder aanpassing aan de objecten
2. Replicatie, het migreren van objecten naar een ander medium zonder aanpassing aan de objecten
3. Repackaging, het migreren van objecten naar een ander informatiepakket zonder aanpassing aan de content
4. Transformatie, het migreren van objecten naar een ander bestandsformaat inclusief aanpassing van de Content Information binnen het bestandsformaat

De eerste drie vormen van migratie zijn belangrijk bij zowel bit-preservering als functionele preservering. De laatste vorm komt uitsluitend voor bij functionele preservering. In dit geval veranderen de bestanden namelijk waardoor de checksum ook anders wordt.

De laatste vorm van migratie is wat we hier bedoelen met formaatmigratie. Deze vorm kan bestaan uit omkeerbare en onomkeerbare transformatie.[33] In het eerste geval kan het nieuwe formaat weer in het oude formaat omgezet worden zonder verlies van data. In het tweede geval kan de oorspronkelijke data niet meer precies gereconstrueerd worden en moet aangetoond worden dat een representatie van de oorspronkelijke data behouden is aan de hand van *significante karakteristieken*.

Onder significante karakteristieken verstaan we de bestanddelen of deelaspecten van een object zonder welke het object niet of niet geheel authentiek is. Deze leggen we in eerste instantie vast op het niveau van de intellectuele entiteit in abstracte, inhoudelijke termen. Een migratieplan moet een verantwoording bevatten hoe de significante karakteristieken vertaald zijn naar technische eigenschappen binnen het nieuwe formaat. Deze informatie wordt ook duurzaam vastgelegd in collectieprofielen.

De eerste twee vormen van migratie worden voor de collectie op regelmatige basis gedaan. Aangezien deze acties geen verandering van de objecten als resultaat opleveren is er voor gekozen om deze niet vast te leggen als event-geschiedenis in de metadata van het object. Wel is natuurlijk van belang om aan te tonen dat de data nog integer is, hetgeen gebeurt aan

[33] "Reference Model For An Open Archival Information System (OAIS)", https://public.ccsds.org/pubs/650x0m2.pdf , pagina 5-4ff

de hand van integriteitscontrole. Met integriteit worden dus alle integriteitscontroles bedoeld en niet alleen de checksumcontrole. Verder worden de migratieacties inclusief planning en procedure vastgelegd in documentatie die ook duurzaam bewaard wordt.

De derde vorm van migratie wordt momenteel ook uitgevoerd en deze wordt wel vastgelegd als event-metadata, bijvoorbeeld als event van het type ‘creation’ bij het nieuw gemaakte object. Ook bij migraties kan dit spelen wanneer de structuur van de AIP of de manier waarop deze structuur vastgelegd is in een systeem wijzigt. In zulke gevallen wordt in een preserveringsactieplan vastgelegd hoe de structuur vernieuwd gaat worden, of er mogelijk nog verschillende scenario’s zijn en hoe de binnen de nieuwe structuur rekening is gehouden met de noodzakelijke randvoorwaarden wat betreft integriteit, authenticiteit en duurzame toegankelijkheid.
De vierde vorm van migratie gaat spelen wanneer uit duurzaamheidsoverwegingen wordt gekozen bepaalde formaten anders op te slaan. Een voorbeeld hiervan is het uitgepakt opslaan van oorspronkelijk ingepakte formaten. Het gaat hierbij om een omkeerbare transformatie. Hiervan zal ook event-metadata worden opgeslagen en een nieuwe checksum worden gegenereerd om de bit-integriteit in de toekomst te kunnen controleren van de nieuwe bestanden.

Kennisniveaus

Zoals hierboven genoemd is de ambitie om in de toekomst onze kennis van welke formaten in de collectie beschikbaar zijn te vergroten om zo risico’s op het gebied van toegankelijkheid te kunnen ondervangen door middel van functionele preservering. Uitgangspunt hierbij is dat er uiteindelijk technische informatie beschikbaar is over alle formaten in de collectie en dat er voldoende kennis in huis is over deze formaten om deze aan de hand van verschillende preserveringsstrategieën voor de lange termijn toegankelijk te houden. Het opdoen van kennis en het analyseren van objecten is een groeimodel waarbij formaten een niveau toegekend krijgen. Met dit groeimodel wordt wel op korte termijn gestart door in kaart te brengen welke bestandsformaten we op dit moment bewaren en wat het niveau van bewaren is. Daarnaast worden goede en slechte voorbeelden bijgehouden van bestandsformaten.
Het niveau van kennis zal dan uiteindelijk gaan van opgeslagen (niveau 1) naar gekend (niveau 3). De kennisniveaus zoals we die hier hanteren betreffen dus een momentopname: uiteindelijk is het doel om objecten te doen stijgen in het niveau, maar het zou kunnen dat sommige bestandsformaten nooit tot het hoogste niveau komen.

De classificatie is als volgt en wordt toegekend per formaat:

- Opgeslagen: opgeslagen zonder bestandsformaatidentificatie
- Geïdentificeerd: opgeslagen met basis-bestandsformaatidentificatie zoals MIME-type
- Gekend: opgeslagen met uitgebreide technische metadata en preservation watch

Zie voor een meer gedetailleerde beschrijving het document Richtlijnen bestandsformaten[34].

[34] [locatie volgt nog]

Alleen wanneer een bestandsformaat gekend is kan er sprake zijn van functionele preservering. Het preserveringsniveau van een collectie is dus gekoppeld aan het kennisniveau van de formaten binnen die collectie.

Expertise

Doordat we van bit preservation naar functionele preservering gaan is er extra kennis nodig over digitale duurzaamheid. Sommige kennis is wel theoretisch gekend maar moet geoperationaliseerd worden in een werkbare en bruikbare implementatie.[35] Hiervoor moeten een aantal acties genomen worden, zoals actief deelnemen aan de duurzaamheidsgemeenschap of aan gemeenschappen rondom specifieke oplossingen, bijvoorbeeld rondom de tools die we gebruiken voor kwaliteitscontrole of bestandsidentificatie. Door actief bij te dragen middels kennisoverdracht binnen deze gemeenschappen komen we tot nieuwe inzichten en kunnen we dit ook als een platform gebruiken om aandacht te vragen voor digitale duurzaamheid.
Een andere manier om ervaring op te doen is om een betere testinfrastructuur op te zetten met bijvoorbeeld probleembestanden die gebruikt kunnen worden om tools te testen en om regressietesten uit te voeren in updates van tools. Zo worden mogelijke problemen voor de verwerking en de duurzame bewaring snel gedetecteerd en kan er ervaring opgedaan worden met nieuwe tools waarmee het kennisniveau over de collecties verdiept kan worden.

Collectieprofielen

Veel van de informatie zoals hierboven beschreven, te weten de preserveringsniveaus, strategieën, kennisniveaus, significante karakteristieken en speciale context-informatie, is steeds onderhevig aan verandering en verschilt vaak per collectie. Het geldt in veel gevallen wel voor grote groepen van objecten zodat het geen nut heeft om dit per object vast te leggen. Om deze reden kiezen we er bewust voor om dergelijke informatie niet vast te leggen als metadata binnen de AIP maar dit te verwerken in overkoepelende collectieprofielen. Zoals hierboven genoemd zullen deze collectieprofielen ook van belang zijn voor de informatie-integriteit van de objecten doordat hierbinnen informatie vastgelegd kan worden die noodzakelijk is om de objecten binnen een collectie te begrijpen. Het dient daarmee ook het aspect ‘interpreteerbaar’ binnen het concept van duurzame toegankelijkheid. Aan de hand van de collectieprofielen kan ook duidelijk gemaakt worden welke objecten tot de AIP behoren zodat hier voor interne en externe gebruikers duidelijkheid over is.

Preserveringsplanning

Binnen preserveringsplanning komen ten slotte alle bovengenoemde begrippen samen: het doel is om functionele preservering toe te passen en de toegankelijkheid van de objecten voor de lange termijn te garanderen door het uitvoeren van een preserveringsstrategie waarbij zorg gedragen wordt dat integriteit en authenticiteit behouden blijft. Hierbij moet dan ook verantwoord worden hoe de significante karakteristieken bij het toepassen van de preserveringsstrategie behouden zijn gebleven. Randvoorwaardelijk hierbij is dat er input is

[35] Dit is een invulling van ISO-16363 eis 3.2.1.3

vanuit de verschillende monitoring mechanismes en dat het hoogste kennisniveau is toegekend aan de betreffende objecten.
Een preserveringsactieplan moet vervolgens opgesteld worden dat bovengenoemde zaken beschrijft ter verantwoording met daarbij de specifieke preserveringsactie die uitgevoerd moet worden. Dit document dient dan als bewijsstuk dat bijdraagt aan de authenticiteit van het object en tegelijkertijd als verslaglegging dat preserveringsacties volgens vaste procedures worden uitgevoerd.[36] Als input voor dit preserveringsactieplan zal technische metadata gebruikt worden. Onder andere op basis van deze informatie worden duurzaamheidsrisico's in kaart gebracht en eventuele maatregelen geformuleerd in het preserveringsactieplan.

Informatieveiligheid

Om te laten zien hoe de verschillende aspecten van informatieveiligheid zijn belicht binnen het duurzaamheidsbeleid maken we gebruik van een categorisering van mogelijke bedreigingen[37] en verwijzen we in welke stukken deze aan de orde komen of stellen we voor wat er toegevoegd moet worden om nog niet-geïdentificeerde risico's te vermijden.

De basisprincipes rondom security van de IT-infrastructuur zoals opgenomen in het Securitybeleid worden als randvoorwaardelijk gezien voor de garantie tegen bedreigingen zoals het falen van media, software, hardware, menselijk handelen en interne en externe aanvallen.[38] De serverruimtes moeten beveiligd worden, zodat onbevoegden geen fysieke toegang hebben tot de data.
Incidenten zoals natuurlijke rampen zijn beschreven in het bedrijfsnoodplan en het collectiehulpverleningsplan. Daarnaast moet de infrastructuur ook ingericht worden zodat risico's gespreid worden. Het uitgangspunt moet zijn dat er geen single points of failures zijn in de infrastructuur of dat er plannen zijn om in het geval van dienstonderbrekingen een oplossing te bieden, maar dat er zeker geen dataverlies optreedt door single points of failure. Risico's met betrekking tot veroudering van hardware en software worden onder controle gehouden middels het opzetten van het proces voor preserveringsplanning.
Tenslotte moeten uiteindelijk ook economische en organisatorische risico's belegd gaan worden in de organisatie in een meerjarenbeleid. Een successieplan zou hier ook deel van moeten uitmaken. Dit is een plan voor overdracht van de collectie als er onvoldoende middelen aanwezig zijn om een deel of de gehele collectie langer duurzaam te kunnen bewaren.
Zoals uit deze opsomming al blijkt is informatieveiligheid veel meer dan slechts een IT-aangelegenheid en is het van belang dat hier ook naar gekeken wordt vanuit het oogpunt van het langdurig behoud van de digitale collectie.
Om informatieveiligheid specifiek voor de digitale collectie te kunnen waarborgen moeten er naast het securitybeleid aanvullende zaken vastgelegd worden. Deels kan dit door duurzaamheidseisen toe te voegen aan bestaande documenten, deels zal dit gebeuren door het opstellen van nieuwe documenten. Zo moet er bijvoorbeeld aangetoond kunnen worden dat personeel op de hoogte is van welke acties in het dagelijks werk impact kunnen hebben

[36] Dit is een invulling van ISO-16363 eis 4.3.4
[37] Ontleend aan http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html
[38] Ibid.:Media failure, Software failure, Hardware failure, Operator error, Internal attack en External attack

op de duurzaamheid van de digitale collectie. Dit zal op operationeel niveau vorm krijgen door het opstellen van een gedragscode en door het verwerken van aspecten rondom duurzaamheid in noodplannen. Bij migraties moet er bijvoorbeeld voor gezorgd worden dat opslagmedia volledig gewist wordt zodat objecten niet beschikbaar komen op een onvoorziene manier[39]. Dit zal vastgelegd worden als onderdeel van migratieplannen. Daarnaast moet er ook een document voor disaster recovery opgesteld worden waarin duidelijk staat beschreven wat er moet gebeuren wanneer objecten of informatie in gevaar zijn en welke procedures er gevolgd moeten worden om de objecten te herstellen. Het is daarbij ook belangrijk dat de procedures regelmatig getest zijn zodat er zekerheid is dat objecten daadwerkelijk hersteld kunnen worden op basis van de beschreven maatregelen. Ook binnen alle projecten waar dit van toepassing is moet disaster recovery een vast onderdeel uit gaan maken van de testprocedure. Ook hier is het niet alleen van belang dat er voorzieningen zijn getroffen voor de backup zelf maar dat de effectiviteit van de backupprocedure ook wordt geverifieerd door middel van een restore, bijvoorbeeld naar een acceptatieomgeving.
Tenslotte moet de informatieveiligheid in sommige gevallen ook verhoogd worden door bepaalde initiatieven tot een proces te maken. Zo is er momenteel al een uitwerking voor het opstellen van een dataclassificatie door middel waarvan inzicht wordt gegeven in het benodigde beveiligingsniveau voor data en de bijbehorende systemen. Ook zijn er sessies geweest om risico's rondom het Digitaal Magazijn in kaart te brengen en deze in te schatten op basis van kans en impact. Voor informatieveiligheid is het van belang dat onderkende risico's die hieruit naar voren komen ook daadwerkelijk geïmplementeerd worden op basis van de resultaten van de dataclassificatie en ook op regelmatige basis opnieuw worden geëvalueerd zodat aantoonbaar wordt dat gesignaleerde risico's worden opgevolgd. Hetzelfde geldt voor de eerdergenoemde gedragscode voor het betrouwbaar omgaan met collectiedata, deze moet niet alleen opgesteld worden maar ook controleerbaar worden gemaakt en geïmplementeerd binnen actuele processen zodat aantoonbaar is dat deze daadwerkelijk gehandhaafd wordt.

Rollen en verantwoordelijkheden

Er moet gezorgd worden dat de functieprofielen de rol en de verantwoordelijkheid bevatten die iemand heeft in het bewaren van de digitale collectie voor alle functies waarbij mensen te maken hebben met data uit de collectie. Ook op het gebied van personeelsbeleid is het belangrijk dat er nog beter inzichtelijk wordt gemaakt hoe de organisatie zorgt dat er genoeg financiën zijn voor het behouden van inhoudelijke expertise rondom duurzaamheid en opleidingsmogelijkheden voor het vergroten en vernieuwen hiervan.[40] Dit is in de inleiding eveneens genoemd als een van de randvoorwaardelijke zaken voor duurzaamheid. Om aan te tonen hoe er binnen de organisatie voldoende rollen aanwezig zijn om duurzaamheid vorm te geven is gebruik gemaakt van het SHAMAN-model.[41]

---

[39] Dit is ook vereist vanuit de Richtlijn Privacy en Security (https://plein.kb.nl/thoughts/9360) waarin verwezen wordt naar DIN 66399 of HMG Infosec Standard 5 of vergelijkbaar bij het wissen van media.
[40] Ibid. 3.2.1 en onderliggende eisen, Blz. 3-3 – 3-5
[41] Met name Operator error en Internal attack zoals beschreven in http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html.

Het doel is om uiteindelijk rollen en verantwoordelijkheden van alle betrokkenen ook vast te leggen in documentatie zodat op basis daarvan toegangsrechten binnen de verschillende systemen van de architectuur toegekend kunnen worden.[42]

Actiepunten

Op basis van bovengenoemde uitwerking volgt hieronder een lijst met concrete actiepunten van eerder genoemde acties die in de nabije toekomst uitgevoerd moeten gaan worden om de duurzaamheid van de collectie en de processen daaromheen te verbeteren.

1. Opstellen van collectieprofielen met daarin informatie over integriteit, authenticiteit, duurzame toegankelijkheid, preserveringsstrategie en significant properties
2. Opstellen IP-overzichtsstructuur
3. Implementeren van uitgangspunten minimale ingest
4. Uitwerken aansluitproces volgens PAIMAS
5. Opzetten monitoring designated communities
6. Formaliseren/uitbreiden preservation watch
7. Integratie contractbeheer
8. Uitbreiden bestandsformaatkennis door:
    a. deelnemen aan communities
    b. analyseren van de collectie
    c. verzamelen van goede en slechte voorbeelden
9. Vastleggen kennisniveaus voor bestandsformaten en een actieplan maken om het kennisniveau te laten stijgen.
10. Formaliseren Preservation planning door het maken van preserveringsactieplannen
11. Toevoegen van duurzaamheidseisen aan het bedrijfsnoodplan en het collectiehulpverleningsplan
12. Opstellen successieplan
13. Inrichten van actueel te houden en controleerbare processen rondom:
    a. Digitale duurzaamheid
    b. Dataclassificatie
    c. Risicomanagement
    d. Disaster recovery
    e. Gedragscode
14. Opstellen rollen en verantwoordelijkheden ten behoeve van toegangsrechten

[42] Dit is een invulling van ISO-16363 eis 5.2.3