



Justitiële Informatiedienst
Ministerie van Justitie en Veiligheid

Preservatiestrategie CDD+

Versie 1.0

Datum	29 augustus 2025
Status	Definitief

Inhoud

	Voorwoord 5
	Inleiding 7
1.1	Informatieobject en verschijningsvorm 7
1.2	Scope van dit document 8
1.3	Samenhang visie, beleid en implementatie 8
2	Doelgroepen 11
3	Betrokken medewerkers 13
3.1	Expert informatiehuishouding 13
3.2	Business informatieanalist 13
3.3	Specialist archiefbeheer 13
3.4	Bestandsspecialist 13
3.5	Applicatie architect en BusDevops engineer 14
4	Visie 15
4.1	Preservatie is een integraal onderdeel van het CDD+ 15
4.2	Centrale strategie: Conversie 15
4.3	Als het niet kan zoals het moet ... 16
4.4	Omgaan met foutieve invoer 16
5	Preservatiebeleid 18
5.1	Metagegevens 20
5.2	Integriteit 20
5.2.1	Redundantie 20
5.2.2	Backups 21
5.2.3	Fixity checks 21
5.3	Bruikbaarheid 22
5.3.1	Detectie van bestandsformaat 22
5.3.2	Validatie en problematische eigenschappen 23
5.3.3	Conversie van bijlagen en embedded bestanden 24
5.4	Authenticiteit 25
5.4.1	Detectie en validatie digitale handtekeningen 25
	Bijlage: Indeling van bestandsformaten 26

Voorwoord

De Justitiële Informatiedienst (Justid) heeft als missie om bij te dragen aan een veilige en rechtvaardige samenleving. Dit doet Justid door belanghebbenden te voorzien van cruciale informatie uit het domein van Justitie, Veiligheid, Asiel en Migratie.¹ Dit betekent onder meer dat documenten² gedurende hun afgesproken bewaartermijn beschikbaar, betrouwbaar en bruikbaar moeten zijn voor deze belanghebbenden. Ook als die bewaartermijn meerdere decennia omvat.

Justid ontwikkelt en beheert voor dit doeleinde het Centraal Digitaal Depot (CDD+) als gemeenschappelijke basisdienst voor de ministeries van Justitie en Veiligheid en Asiel en Migratie. Dit systeem is ingericht op basis van het referentiemodel Open Archival Information System (OAIS). In het CDD+ kunnen documenten duurzaam worden bewaard, zodat ze gedurende hun hele levensduur vindbaar, beschikbaar, leesbaar (waarneembaar), interpreteerbaar, betrouwbaar en toekomstbestendig blijven. Oftewel: duurzaam toegankelijk.

De voornaamste doelgroep waar het CDD+ zich op richt zijn daartoe gerechtigde ambtenaren die werken in bovengenoemde domeinen. Daarnaast is er oog voor toekomstige doelgroepen van informatie in het CDD+: burgers en onderzoekers. Zij kunnen permanent te bewaren documenten op termijn inzien bij een archiefdienst zoals het Nationaal Archief.

Dit document beschrijft de preservatiestrategie en het preservatiebeleid van de Digitaliserings- en Archiveringsdienst van Justid.

¹ <https://www.justid.nl/over-ons>.

² In deze preservatiestrategie worden de termen 'document' en 'informatieobject' allebei gebruikt. Met deze termen wordt hetzelfde bedoeld.

Inleiding

Zoals vermeld in de grondslagen³ van het CDD+ is de preservatiestrategie onderdeel van de "dienst" CDD+. De preservatiestrategie heeft tot doel de kwaliteit te borgen van de informatieobjecten (documenten) die in het CDD+ zijn opgenomen.

NEN-ISO 15489-1:2016 (waarnaar verwezen wordt in de Memorie van Toelichting bij de Archiefregeling) geeft een viertal kwaliteitseisen die aan informatieobjecten worden gesteld:

- authenticiteit is een kwaliteitseis aan de (bevoegde) actor, die op een bepaald tijdstip een bepaald informatieobject heeft ontvangen of opemaakt.
- betrouwbaarheid is een eis aan de vastlegging van de handeling, of deze volledig en nauwkeurig wordt vastgelegd. Hoe gedetailleerder de procedure, hoe betrouwbaarder het informatieobject is. Bij informatieobjecten heeft dit een nauwe relatie met de authenticiteit van diegene die de handeling vastlegt/vast moet leggen (de bevoegde actor);
- integriteit is een eis aan de inhoud. De inhoud mag niet onbevoegd of onbedoeld worden gewijzigd;
- bruikbaarheid is een eis aan de interpreteerbaarheid van de verschijningsvorm. Verschijningsvormen die niet bruikbaar zijn, zijn waardeloos in ieder (werk)proces.

Deze strategie beschrijft maatregelen die in het CDD+ worden genomen om aan deze eisen te voldoen. De meeste omvangrijke maatregelen zijn nodig om de bruikbaarheid te borgen. Deze is bij digitale informatieobjecten minder vanzelfsprekend dan bij papieren objecten, en maatregelen die nodig zijn om de bruikbaarheid te borgen, kunnen de integriteit van de inhoud aan tasten.

1.1 Informatieobject en verschijningsvorm

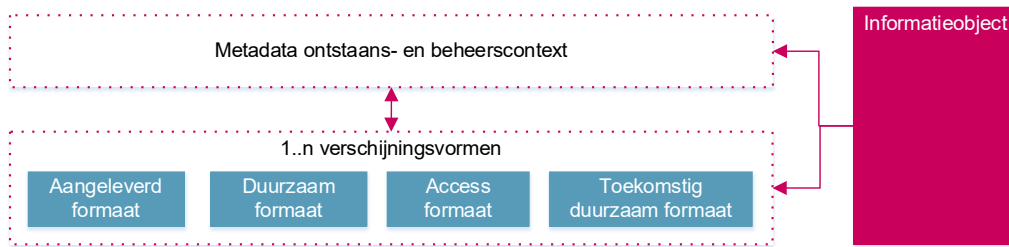
In dit stuk hebben we het over "informatieobjecten" (documenten) en "verschijningsvormen".

Een informatieobject is het opgeslagen document met al zijn verschijningsvormen plus metadata. Dit betreft metadata uit de ontstaans- en beheercontext rondom het informatieobject. Elk informatieobject krijgt een uniek identificatienummer: een *Digital Object Identifier* (DOI).

Een verschijningsvorm is een bitstream die op een bepaalde wijze in een bestandsformaat is verpakt. Denk aan een DOCX, een PDF, een JPEG of een MP3. De verschijningsvormen van een informatieobject kunnen bestaan uit een aangeleverd formaat en – indien mogelijk – een naar een duurzaam formaat geconverteerde verschijningsvorm.

In voorbereiding is functionaliteit om in bepaalde situaties ook aan de *access* kant op verzoek een conversie te doen naar een formaat waar de vragende partij behoefte aan heeft. Indien die situatie zich voordoet, wordt ook deze verschijningsvorm bewaard als onderdeel van het informatieobject in het CDD+.

³ Digitalisering en Archiveringsdienst, Nota Grondslagen e-depotvoorziening, v3.0, 24 maart 2023.



1.2 Scope van dit document

De informatie in dit document bestaat uit twee delen:

- De *visie*: dit deel beschrijft de kaders waar de preservatiemaatregelen aan moeten voldoen, en een aantal principiële keuzes die zijn gemaakt over hoe de preservatie in het CDD+ is vormgegeven.
- Het *beleid*: dit deel beschrijft de concrete maatregelen die getroffen moeten worden. Deze maatregelen zijn gebaseerd op de visie en *best practices* binnen de archivaliek. Ze worden waar nodig toegespitst op de binnen het CDD+ gangbare vormen van informatieobjecten.

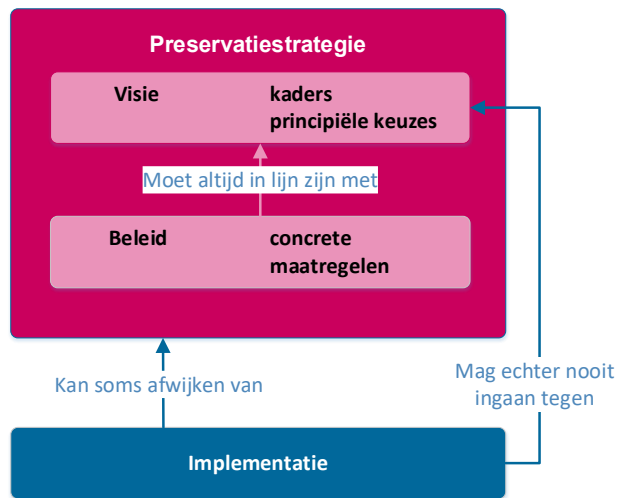
Naast de visie en het beleid bestaat er nog een derde laag. Dit document beschrijft dit derde deel niet:

- De *implementatie van het beleid*: dit zijn de keuzes die worden gemaakt om de in het beleid beschreven maatregelen uit te voeren. Dit gaat onder meer om keuzes voor software, software-instellingen, algoritmes, en procedures.

1.3 Samenhang visie, beleid en implementatie

Het beleid moet altijd in lijn zijn met de visie. Echter, er komen doorlopend nieuwe bestandsformaten bij, en nieuwe mogelijkheden om met die bestandsformaten om te gaan. Daardoor zal de implementatie van het beleid vaak achterlopen op het beleid.

Idealiter is de implementatie van het beleid weer in lijn met het beleid. In de praktijk zal dit echter vaak niet haalbaar zijn. Door beperkingen in de beschikbare software, budgetten, en capaciteit, zal het vaak zo zijn dat de implementatie van het beleid niet alle elementen van het beleid volledig invult. Hoewel de implementatie van het beleid kan afwijken van het beleid, mag hij niet ingaan tegen de visie.



De visie ligt grotendeels vast, en zal doorgaans alleen wijzigen door ingrijpende veranderingen in de omgeving van het CDD+, zoals een wetwijziging. Wel is het mogelijk dat de tekst van de visie wordt bijgesteld om de intentie beter of nauwkeuriger over te brengen. De visie wordt onderhouden door de experts informatiehuishouding en vastgesteld door het MT van de Digitalisering en Archiveringsdienst (DAD). Elk jaar wordt beoordeeld of er aanleiding is om de visie aan te passen.

Het beleid wordt aangepast en uitgebreid telkens als blijkt dat het de informatieobjecten die worden aangeboden niet voldoende dekt – dus als er nieuwe formaten worden toegevoegd, of documenten met eigenschappen die niet goed samengaan met het bestaande beleid⁴. Het beleid wordt onderhouden door de experts informatiehuishouding en vastgesteld door het MT. Het preservatiebeleid is een integraal onderdeel van het CDD+, en alle medewerkers van de archiveringsdienst kunnen tegen de beperkingen van het beleid aanlopen, en deze aangeven bij de experts informatiehuishouding. De bestandsspecialisten en de specialisten archiefbeheer hebben hier een expliciete rol, en worden geacht bepaalde aspecten van het beleid te monitoren en daarover te rapporteren. Dit wordt in het beleid zelf nader uitgewerkt.

De implementatie van het beleid ligt bij verschillende teams binnen de archiveringsdienst. Zie hiervoor hoofdstuk 3.

⁴ Merk op dat een uitbreiding van het beleid niet altijd zal betekenen dat de nieuwe formaten of eigenschappen volledig ondersteund gaan worden; het kan ook betekenen dat expliciet gemaakt wordt dat ze niet of niet volledig ondersteund zullen worden.

2 Doelgroepen

De doelgroep of “designated community”, die Justid met het CDD+ wil bedienen bestaat uit de ambtenaren die werken bij een van de uitvoeringsorganisaties in het domein van Justitie en Veiligheid en Asiel en Migratie. Daarnaast is er aandacht voor doelgroepen in de toekomst: geïnteresseerde burgers en onderzoekers die recht hebben op toegang tot naar een archiefdienst overgebrachte en openbaar geworden archieven.

Doelgroep ambtenaren

Bij ingebruikname van het CDD+ krijgen ambtenaren gedeeltelijk of volledig toegang tot de applicatie en tot informatie, conform geldende wet- en regelgeving. Documenten worden beschikbaar gesteld op basis van het ‘need-to-know’ principe. Dit komt terug in de missie van Justid. De doelgroep van ambtenaren bestaat uit twee subdoelgroepen: informatiespecialisten die de bedrijfsvoering ondersteunen en ambtenaren die bedrijfsprocessen uitvoeren.

Informatiespecialisten

Deze subdoelgroep bestaat uit functionarissen die werkzaam zijn in archief- en informatiemanagement. Zij voeren in het CDD+ archiefbeheertaken uit voor het archief / de archieven van de desbetreffende organisatie waar zij voor werken. In dat kader moeten zij de documenten en bijbehorende metadata kunnen raadplegen en zo nodig herstelacties kunnen uitvoeren. Ook zijn zij verantwoordelijk voor het vernietigen of overbrengen van documenten wanneer de bewaartermijn verstreken is.

Knowledge Base:

De informatiespecialisten vormen een belangrijke schakel tussen de ambtenaren die het CDD+ gebruiken als een van de bronsystemen waarin zij informatie kunnen zoeken en raadplegen.

- Zij kennen de primaire werkprocessen en werkwijze van hun eigen organisatie goed.
- Daarnaast hebben ze kennis van ondersteunende processen, waaronder het proces van dossiervorming.
- Alle klanten die op het CDD+ aangesloten zijn hebben functioneel beheerders, die kennis hebben van gebruikersbeheer (accounts, autorisaties, attributen etcetera), van berichtenverkeer en van metadata. Bij een aantal organisaties is er een archivaris of recordmanager werkzaam. Beheerders met kennis van metadatastandaarden, ordeningsstructuren en van waardering en selectie zijn echter niet altijd (voldoende) aanwezig. Hierover kan Justid zo nodig adviseren.
- Voor gebruikers die inloggen via de CDD+ GUI zijn op een aparte Help pagina Quick Reference Cards (QRC's) beschikbaar. Ook zijn er handleidingen voor vernietigen op basis van een selectielijst en overbrengen van informatieobjecten naar het Nationaal Archief.

Uitvoerend ambtenaren

Deze subdoelgroep bestaat uit ambtenaren die, wanneer zij daartoe gerechtigd zijn, het CDD+ gebruiken als een van de systemen waarin zij informatie kunnen zoeken, opvragen en raadplegen. Dit doen zij in het kader van de werkprocessen en taken die ze uitvoeren. Tot deze groep behoren bijvoorbeeld:

- beslismedewerkers van de Immigratie- en Naturalisatiedienst (IND), die besluiten over het inwilligen of afwijzen van verblijfsvergunningen aan vreemdelingen;
- regievoerders van de Dienst Terugkeer en Vertrek (DT&V) die, wanneer een vreemdeling Nederland moet verlaten, het vertrekproces regisseren;
- medewerkers van de Dienst Justitiële Inrichtingen, die informatie moeten kunnen raadplegen over justitiabelen die in bewaring zijn gesteld bij de penitentiaire inrichting waar zij werken;
- daarnaast kunnen bovengenoemde beheerders ook bepaalde informatie uit hun CDD+ archief beschikbaarstellen aan derden, bijvoorbeeld advocaten.

Knowledge Base:

Het zijn vaak hoogopgeleide personen die tot deze doelgroep behoren.

- Naast de Nederlandse taal zijn ze bekend met ambtelijk jargon, in het bijzonder jargon dat gangbaar is in het domein van Justitie en Veiligheid.
- Ze hebben veel kennis van de eigen primaire werkprocessen, ambtelijke procedures en die van ketenpartners.
- Ze weten wat voor soort informatie ze nodig hebben om zaken af te handelen.

Doelgroep burgers en onderzoekers

Het beleid voor het CDD+ is niet alleen gericht op de gebruikers in het hier en nu, die informatie nodig hebben om hun werk adequaat te kunnen doen. Bij een aansluiting van een organisatie op het CDD+ wordt de organisatorische en juridische context zo goed mogelijk in kaart gebracht. Ook wordt ernaar gestreefd zoveel mogelijk metadata te verzamelen over de ontstaans- en beheercontext van documenten.

“Nu” is die context beter bekend, dan op een later tijdstip. De kans is groot dat op het moment dat documenten moeten worden overgebracht de archiefvormende organisatie een of meer reorganisaties heeft doorgemaakt – of nog lastiger – is gefuseerd of verdwenen. Het zou dan veel moeite, tijd en geld kosten om de ontstaans- en beheercontext te reconstrueren, met alle gevolgen van dien voor integriteit, bruikbaarheid, authenticiteit en betrouwbaarheid van informatie.

Op die manier is er aandacht voor toekomstige doelgroepen van de archieven in het CDD+: geïnteresseerde burgers en onderzoekers, die vragen kunnen hebben als:

- Wat was het beleid van een organisatie?
- Wie wist wat, wanneer?
- Hoeveel zaken werden er in een bepaalde periode afgehandeld?
- Hoe werd invulling gegeven aan een bepaald mandaat?

3 Betrokken medewerkers

Preservatie begint al bij de creatie of ontvangst van een informatieobject. Dat betekent dat een belangrijk deel van de werkzaamheden bij de ketenpartners ligt. Het is echter ook een hoofdtaak van het CDD+. Dit betekent dat alle medewerkers van de DAD die zich bezighouden met het CDD+ ook oog moeten hebben voor preservatie, en dat preservatie invloed zal hebben op hun werkzaamheden. Ze zullen verantwoordelijk zijn voor het toepassen van (een deel van) het preservatiebeleid, en er wordt van hen verwacht dat ze problemen bij die toepassing signaleren, zodat het beleid en de implementatie ervan kan worden bijgesteld.

Er is een aantal medewerkers die een meer nadrukkelijke rol hebben.

3.1 Expert informatiehuishouding

Experts informatiehuishouding zijn verantwoordelijk voor het opstellen van de preservatiestrategie – dit document. Dit betekent dat zij aan de hand van wet- en regelgeving, gangbare normen, technische mogelijkheden, *best practices*, en de situatie bij de ketenpartners vaststellen welke preservatiemaatregelen in het CDD+ zouden moeten worden genomen. Vervolgens bepalen zij ook op welke wijze daarvan af wordt geweken, als blijkt dat de ideaalsituatie niet haalbaar is, als gevolg van technische, budgettaire, of andere beperkingen. Het bepalen van welke bestandsformaten worden ondersteund en worden ingezet bij preservatie⁵ is een onderdeel van de preservation planning uit het OAIS-model, die ook wordt uitgevoerd door de experts informatiehuishouding. Deze indeling maakt ook deel uit van deze preservatiestrategie.

3.2 Business informatieanalist

De business informatieanalist stelt de business requirements op voor benodigde functionaliteit in het CDD+, op basis van de visie en het beleid, en beheert deze in de requirements repository. Ook staat hij in nauw contact met het ontwikkelteam voor de nadere uitwerking van de business requirements.

3.3 Specialist archiefbeheer

De specialisten archiefbeheer houden zich bezig met de wijze waarop informatieobjecten worden beschreven in het CDD+, en zijn dan ook verantwoordelijk voor de keuzes ten aanzien van de archivalische beschrijvingen.

De specialisten archiefbeheer monitoren ook de bruikbaarheid van de gekozen preservatieformaten. Als het maatschappelijk draagvlak (waaronder de software ondersteuning) van een formaat dreigt weg te vallen, dan rapporteren zij dat, zodat het CDD+ tijdig kan overstappen op andere formaten.

3.4 Bestandsspecialist

De bestandsspecialisten (een rol die doorgaans vervuld zal worden door functioneel beheerders) houden zich bezig met de technische eigenschappen van de bestanden waarmee gewerkt wordt, en de technische eigenschappen van de software waarmee deze bestanden verwerkt worden om invulling te geven aan het preservatiebeleid.

⁵ Dit wordt verder uitgelegd in hoofdstuk 5.

De bestandsspecialisten zijn verantwoordelijk voor: ⁶

- De selectie van de benodigde software, en de instellingen ervan, waarbij ze samenwerken met de experts informatiehuishouding als er afgeweken moet worden van de ideaalsituatie;
- het analyseren en (mits mogelijk) oplossen van problemen bij conversie;
- het opbouwen en onderhouden van een testset;
- tijdens het vooronderzoek voor een aansluiting van een klant op het CDD+ nagaan of documenten die worden geproduceerd bij die klant voldoen aan het bestandsformaat;
- doorontwikkeling van bestaande of te verwachten bestandsformaten signaleren, hierover rapporteren, en – na overleg – desgewenst het life cycle management opnemen.

3.5 Applicatie architect en BusDevops engineer

De applicatie architect en het ontwikkelteam zijn verantwoordelijk voor de implementatie van de maatregelen die niet worden gerealiseerd middels externe software, en maken de keuzes voor de algoritmes die worden ingezet bij de realisatie van deze maatregelen. Indien zij inschatten dat een bepaald onderdeel van het preservatiebeleid niet (volledig) haalbaar is, signaleren zij dit bij experts informatiehuishouding en wordt dit met hen besproken (zie paragraaf 3.1).

⁶ Zie de Memo Procedure duurzame toegankelijkheid v1, 17 oktober 2022 (20221017_Memo_Bestandsspecialist_v1).

4 Visie

4.1 Preservatie is een integraal onderdeel van het CDD+

Preservatie is een integraal onderdeel van het CDD+, en de ontwikkeling en toepassing van de preservatiestrategie is een van de manieren waarop Justid de ketenpartners ontlast. Door het CDD+ te gebruiken, hoeven ketenpartners zelf geen kennis te ontwikkelen van de verschillende bestandstypes, standaarden en de technologie die nodig is om deze duurzaam toegankelijk te maken. DAD neemt ook de monitoring en signalering van ontwikkelingen omtrent deze onderdelen op zich.

Dit betekent ook dat de preservatiestrategie door Justid wordt bepaald. Ketenpartners beslissen niet over de strategie, en hebben niet de keuze om de strategie aan te passen – ook niet voor de eigen archieven. Vanzelfsprekend is het wel zo dat als de preservatiestrategie of de implementatie ervan leidt tot problemen bij de processen van de ketenpartners, Justid stappen zal nemen om deze problemen te verhelpen.

4.2 Centrale strategie: Conversie

In de 'Grondslagen e-depotvoorziening' is de keuze gemaakt voor actieve preservatie. Dat willen zeggen dat er op voorhand al actie wordt ondernomen om (met name) de bruikbaarheid van de informatieobjecten te borgen.

Wat digitale objecten betreft bestaan hiervoor twee gangbare technieken: conversie en emulatie.

Bij conversie worden informatieobjecten omgezet naar een duurzaam bestandsformaat (dat wil zeggen: een formaat waarvan verwacht wordt dat het nog lange tijd bruikbaar blijft). Bij emulatie wordt de software en de hardware die nodig is om de bestanden te openen, nagebootst met behulp van emulatiesoftware. Dit zorgt ervoor dat verouderde bestandsformaten, die in een moderne omgeving niet meer bruikbaar zijn, nog wel geopend kunnen worden in de geëmuleerde omgeving. Eigenschappen en functionaliteiten van een bestand, die bij conversie misschien (gedeeltelijk) verloren gaan, blijven in een geëmuleerde omgeving beschikbaar.

Een nadeel van emulatie is dat de informatieobjecten qua technologie stil staan in de tijd. Bij elke verandering moet je kijken hoe je het oude bestandsformaat nog kunt ondersteunen. Bij conversie beweeg je meer mee met de tijd, en hoef je niet terug te vallen op het oorspronkelijke, oude bestandsformaat, maar kun je eventueel doorbouwen op de meest recente conversie⁷. Ook is bij conversie geen kennis van verouderde systemen en werkwijzen nodig. Een ander nadeel van emulatie is dat het raadplegen van de informatieobjecten gebonden is aan de geëmuleerde omgeving, wat het gebruik van de objecten beperkt.

Emulatie heeft vooral voordelen als de informatieobjecten sterk geïntegreerd zijn met de oorspronkelijk software (of hardware). Voor de meest voorkomende informatieobjecten binnen JenV is dit niet het geval. Dit zijn veelal tekstdocumenten, afbeeldingen, geluidsopnames of beeldopnames, waarvoor veel verschillende ondersteuning bestaat. Daarom is in het verleden de keuze voor

⁷ Indien mogelijk zal er wel voor gekozen worden om te werken met het oorspronkelijke formaat, om te voorkomen dat er een stapeling van conversiefouten optreedt.

preservatie binnen het CDD+ dan ook op conversie gevallen. Inmiddels wordt deze aanpak ook door de wetgeving voorgeschreven⁸. Deze centrale strategie wordt ondersteund door:

- een provenance trail en een functionele trail waarin alle uitgevoerde (preservatie) activiteiten worden vastgelegd en gelogd; ⁹
- metadata;
- maatregelen om problemen met de hardware voor opslag tegen te gaan, door tijdig te migreren naar andere hardware.

4.3 Als het niet kan zoals het moet ...

De grondslagen moedigen "actueel archiveren" aan, hetgeen inhoudt dat een informatieobject direct na creatie wordt ondergebracht in een beheerde omgeving, die de digitale duurzaamheid veiligstelt. In de praktijk zal het echter vaak voorkomen dat een informatieobject pas in een later stadium naar het CDD+ komt, wat het borgen van de duurzaamheid moeilijker maakt.

De grondslagen noemen ook het gebruik van open bestandsformaten. Ook hier kan van worden afgeweken, onder meer omdat ketenpartners informatie ontvangen van derden, die zich niet aan de richtlijnen voor open formaten hoeven te houden.

Met andere woorden: het zal niet altijd mogelijk blijken om het preservatiebeleid perfect toe te passen, omdat de aanlevering van de informatieobjecten niet altijd een ideale vorm zal aannemen. Als dit in de praktijk voorkomt, dan wordt bij voorkeur gekeken of de aanlevering verbeterd kan worden. Echter, als dit niet kan (of niet realistisch is), dan worden de informatieobjecten wel geaccepteerd in het CDD+¹⁰.

Het kan voorkomen dat er op deze manieren objecten in het CDD+ komen waarbij niet geborgd kan worden dat een conversie succesvol zal zijn. In dergelijke situaties wordt de conversie wel uitgevoerd, maar worden de beperkingen in de metadata vastgelegd.

4.4 Omgaan met foutieve invoer

Bij het uitvoeren van de preservatieacties kan het CDD+ tot de conclusie komen dat de eigenschappen van een bestand afwijken van wat de ketenpartner heeft opgegeven. In een dergelijke situatie werkt het CDD+ verder met de eigenschappen zoals het die zelf heeft vastgesteld.

De gegevens die de ketenpartner heeft aangeleverd worden niet overschreven. De eigenschappen die het CDD+ heeft vastgesteld, worden apart opgeslagen. Deze

⁸ Archiefregeling 20xx, art. 2.9, eerste lid: Een document wordt op een zo vroeg mogelijk moment opgeslagen in een duurzaam bestandsformaat.

⁹ Dit maakt deel uit van de "ononderbroken keten van beheer" die in de grondslagen wordt genoemd.

¹⁰ Er zijn een paar uitzonderingen voor bestandsformaten die doorgaans niet gebruikt worden voor informatieobjecten in de context van het CDD+; deze worden in sectie 5 behandeld. Daarnaast kan het voorkomen dat informatieobjecten worden geweigerd op basis van iets anders dan dit preservatiebeleid, zoals veiligheidsoverwegingen.

werkwijze zorgt ervoor de oorspronkelijk aangeleverde informatie beschikbaar blijft voor reconstructies en rapportages.

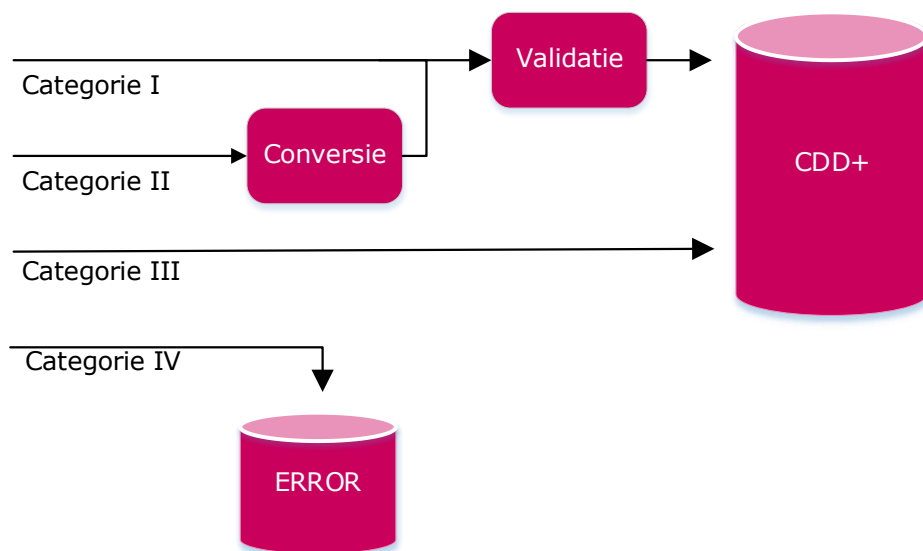
Fouten in de metagegevens kunnen een indicatie zijn voor een fout in de processen van de ketenpartner. De ketenpartner wordt dan ook geïnformeerd over de fouten die worden aangetroffen, zodat processen kunnen worden aangepast om fouten in de toekomst te voorkomen.

5 Preservatiebeleid

Zoals vermeld in sectie 4.1 vormt de conversie van bestanden de basis van de preservatiestrategie. Niet alle bestandsformaten kunnen of hoeven te worden geconverteerd. Er zijn vier verschillende categorieën van bestandsformaten te onderscheiden, afhankelijk van de waarborg voor de mate van bruikbaarheid die wordt gegeven:

- Categorie I: Duurzame bestandsformaten die volledig worden ondersteund.
- Categorie II: Bestandsformaten die worden ondersteund door conversie naar een formaat uit categorie I.
- Categorie III: Bestandsformaten die alleen worden opgeslagen, zonder garantie voor het kunnen openen van de documenten in de toekomst. (De integriteit wordt wel gewaarborgd, zie paragraaf 5.2)
- Categorie IV: Bestandsformaten die niet worden geaccepteerd in het CDD+.

De afhandeling van de verschillende categorieën wordt in onderstaande afbeelding geïllustreerd:¹¹



De indeling van de bestandsformaten is niet statisch, maar zal regelmatig veranderen. Nieuwe formaten zullen worden toegevoegd, en door wijzigingen in de beschikbare software zullen formaten wel of juist niet meer geconverteerd kunnen worden.

Hoewel het bestandsformaat de belangrijkste factor is in hoe een gegevensbestand wordt verwerkt, zijn er ook nog andere bestandseigenschappen die van invloed zijn. Deze kunnen ertoe leiden dat een bestand op een lager niveau ondersteund wordt dan op basis van het bestandsformaat verwacht zou worden. Zie paragrafen 5.3.2 en 5.3.3.

¹¹ Indien nodig wordt na validatie van een categorie I formaat alsnog een conversie van het bestand uitgevoerd, waarna het validatieproces wederom wordt uitgevoerd voordat het bestand in het e-depot wordt opgenomen. Het kan ook voorkomen dat een categorie II formaat niet of niet succesvol geconverteerd kan worden naar een categorie I formaat. Dan wordt het bestand enkel in het aangeleverd formaat opgenomen in het e-depot.

Het streven is om bestandsformaten (voor documenten en archivering) die voorkomen op de lijst met gangbare standaarden of de "Pas toe of leg uit" lijst van het Forum Standaardisatie¹² op te nemen in categorie I of II, zodat alle bestandsformaten die binnen de overheid gangbaar zijn en aangemoedigd worden, gepreserveerd kunnen worden in het CDD+.

Een bestandsformaat kan worden opgenomen in categorie I als de verwachting bestaat dat dit formaat nog lange tijd ondersteund wordt. De belangrijkste voorwaarden hiervoor zijn dat:

- de bestandsindeling openbaar is;
- het formaat alle gegevens bevat die nodig zijn om het document te tonen;
- het formaat breed ondersteund wordt.

Bij voorkeur is het bestandsformaat een open standaard.¹³

Niet elk bestandsformaat dat voldoet aan deze eisen zal worden opgenomen in categorie I; als het eenvoudig geconverteerd kan worden naar een ander categorie I formaat kan er voor worden gekozen om het formaat op te nemen in categorie II.

Een bestandsformaat kan worden opgenomen in categorie II als het betrouwbaar en zonder verlies van informatie kan worden geconverteerd naar een categorie I formaat.

Executables en archive files¹⁴ zijn niet toegestaan in het CDD+. Een uitzondering hierop zijn archive files die embedded zijn in een ander bestand. Bestanden die zijn versleuteld worden ook niet opgenomen in het CDD+. Zonder de sleutel kunnen deze niet worden ingezien en niet worden geconverteerd. Ondersteuning van deze bestanden vergt daarom ook beheer van de sleutel. Dit is complex en (mede daarom) kostbaar. Versleuteling van bestanden voegt ook weinig toe, aangezien het CDD+ al adequaat beveiligd wordt. Een versleuteld bestand valt daarmee in categorie IV, zelfs al heeft het een bestandsformaat dat in een andere categorie geplaatst is.

Andere bestandstypen kunnen worden opgenomen in categorie III indien dit gewenst is. Deze categorie bevat met name:

- Niet-duurzame bestandsformaten die (nog) niet kunnen worden geconverteerd naar een duurzaam formaat.
- Niet-duurzame bestandsformaten die wel kunnen worden opgenomen in categorie II, maar om pragmatische redenen (zoals kosten) daar niet zijn geplaatst.

¹² Het Forum Standaardisatie is een overheidsdienst die de interoperabiliteit en de toepassing van open standaarden binnen de Nederlandse overheid bevordert. Het Forum beheert een lijst met gangbare standaarden, en daarnaast een "Pas toe of leg uit" lijst van open standaarden. (Semi-)overheden zijn verplicht om deze standaarden te gebruiken. Alleen met zwaarwegende redenen mag worden afgeweken van deze lijst.

¹³ De Library of Congress verzamelt kennis over bestandsformaten. Deze inventarisatie van eigenschappen van bestandsformaten vormt voor de DAD een belangrijke bron van informatie. <https://www.loc.gov/preservation/digital/formats/index.html>

¹⁴ Voorbeelden van archive files zijn ZIP en RAR bestanden. Deze bestanden bevatten andere bestanden (eventueel met metadata), maar geen "eigen" inhoud, in tegenstelling tot bijvoorbeeld een e-mail die bestaat uit een tekst en andere bestanden als bijlage.

- Niet-duurzame bestandsformaten die zullen worden opgenomen in categorie II, en in afwachting van de implementatie van de conversie in categorie III zijn geplaatst.

5.1 Metagegevens

Metagegevens beschrijven de context van elk informatieobject, en zijn belangrijk voor het duurzaam toegankelijk houden van de informatie. In de grondslagen van het e-depot zijn metagegevens dan ook benoemd als onderdeel van het CDD+¹⁵.

Hoewel de metagegevens voor preservatie van belang zijn, zijn de te gebruiken metagegevens al elders vastgelegd. In deze strategie worden ze dan ook niet nader beschreven, behalve daar waar ze samenhangen met de maatregelen die genoemd worden.

5.2 Integriteit

Het borgen van de integriteit houdt in dat er maatregelen zijn getroffen om te voorkomen dat er fouten ontstaan in de bestanden die in het CDD+ zijn opgenomen. De wetgeving spreekt in dit kader over een "bit-perfect-exemplaar"¹⁶.

5.2.1 Redundantie

De eerste bescherming tegen hardware fouten, zoals bitrot, is de inzet van een RAID¹⁷ systeem. In een dergelijk systeem wordt de data verdeeld over meerdere harde schijven, en wordt er aan elk blok data een controlegetal toegevoegd, een zogenaamde *parity*. Als een van de harde schijven een beschadiging oploopt, dan klopt de *parity* niet meer, en wordt de fout herkend. Met behulp van de *parity* kan de data worden hersteld. Deze controle wordt elke keer dat het bestand wordt geopend uitgevoerd¹⁸.

Worden er te veel fouten aangetroffen op een schijf, dan wordt aangenomen dat deze defect is, en wordt de informatie verplaatst naar een reserve harde schijf. De defecte schijf wordt binnen een afgesproken tijdsbestek vervangen.

(Naast deze redundantie in opslagmedia is het CDD+ ook als geheel redundant uitgevoerd. Dit betekent dat er twee identieke instanties van het CDD+ bestaan. Als de ene instantie uitvalt, kan de andere instantie het werk overnemen. Deze redundantie borgt de continuïteit van de dienstverlening, en is geen onderdeel van het preservatiebeleid.)

¹⁵ Zie Digitaliserings- en Archiveringsdienst, Nota Grondslagen e-depotvoorziening, v3.0 (maart 2023), aldaar paragraaf 3.2 en 3.4. De wetgever verwijst – in art. 19 Ar – naar NEN-ISO 23081. Voor de Rijksoverheid is van NEN-ISO 23081 de Richtlijn metagegevens Rijksoverheid en het MDTO afgeleid.

¹⁶ Archiefregeling 20xx, artikel 2.10: Het verantwoordelijk overheidsorgaan treft passende maatregelen zodat is gewaarborgd dat van elk document een bit-perfect-exemplaar beschikbaar is.

¹⁷ Redundant Array of Independent Disks.

¹⁸ Dit betekent ook dat een bestand wordt gecontroleerd als het wordt aangesproken om een backup te maken. De backups leiden dus ook tot een periodieke controle.

5.2.2 Backups

De tweede bescherming, zowel tegen hardware fouten als menselijke fouten, bestaat uit backups van de bestanden en de database met metadata. Het CDD+ volgt hierin het algemene backup-beleid van Justid.

5.2.3 Fixity checks

RAID biedt bescherming tegen eenvoudige beschadiging van de hardware. Dit systeem helpt echter niet tegen fouten die ontstaan doordat een bestand verkeerd wordt weggeschreven bij kopiëren of verplaatsen ervan. Daarvoor is een andere methode nodig namelijk gebruik maken van een controlegetal dat wordt berekend over het bestand als geheel, een zogenaamde *hash*. Ook hier geldt dat als het controlegetal is gewijzigd, dit aangeeft dat het bestand is gewijzigd. In dat geval moet er een back-up worden teruggeplaatst. Het controleren van een dergelijke *hash* wordt een *fixity check* genoemd. Fixity is een benaming voor het "onveranderlijk" zijn van een bestand.

Diverse standaarden benoemen het gebruik van fixity checks als (vereist) instrument om te controleren dat de bestanden tijdens de opslag niet veranderd zijn. Het gebruik van fixity metadata wordt beschreven in het OAIS referentiemodel¹⁹, PREMIS²⁰ en in NEN-ISO 16363. De laatstgenoemde standaard geeft aan op welke manieren kan worden aangetoond dat een organisatie actief de integriteit van *archival information packages* (AIPS) monitort:

*Fixity information (e.g., checksums) for each ingested digital object/AIP; logs of fixity checks; documentation of how AIPs and Fixity information are kept separate; documentation of how AIPs and accession registers are kept separate.*²¹

Om te garanderen dat de bestanden onveranderd zijn, worden periodiek alle bestanden – zowel de aangeleverde als geconverteerde verschijningsvormen - gecontroleerd. De frequentie hangt samen met de bewaarperiode van de backup, en moet zodanig gekozen zijn dat fixity check wordt uitgevoerd voordat de laatste backup gemaakt voor de vorige fixity check wordt verwijderd. (Met andere woorden:

¹⁹ CCSDS 650.0-M-3, Reference Model for an Open Archival Information System (OAIS), Magenta Book, december 2024, pagina 4-35 en 4-36:

"Fixity Information provides the data integrity checks or validation/verification keys used to ensure that the particular Content Data Object has not been altered in an undocumented manner. Fixity Information could include special encoding and error detection schemes that are specific to instances of Content Data Objects. Fixity Information does not include the integrity preserving mechanisms provided by the OAIS underlying services, error protection supplied by the media and device drivers used by Archival Storage."

²⁰ Data Dictionary for Preservation Metadata: PREMIS version 3.0, pagina 258:

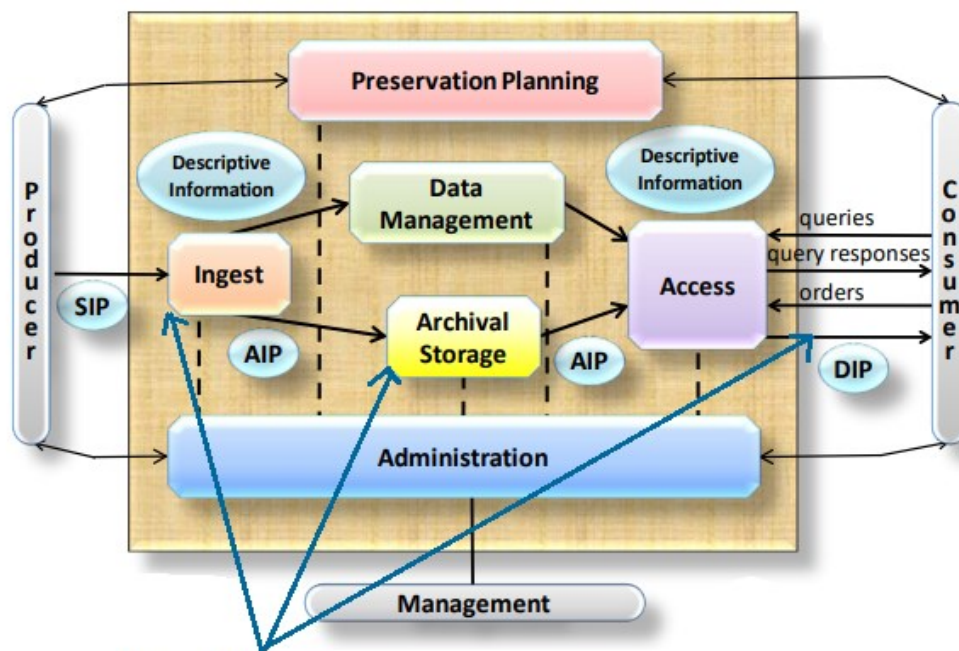
"In the digital domain, hash algorithms that create a message digest can be used to implement a fixity check for an object. If the message digest created by an algorithm at one point is identical to the message digest created by the same algorithm at a later point, this indicates the object did not change during the interim. In fact, it is common to create and test two or more message digests using different algorithms to be certain that an object is fixed."
<http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>

²¹ NEN-ISO 16363:2025 Space data and information transfer systems — Audit and certification of trustworthy digital repositories, pagina 4- 22.

het moet altijd mogelijk zijn om een backup terug te plaatsen van vóór de vorige controle.)

De fixity check vindt plaats door een controlegetal te berekenen, en deze te vergelijken met het controlegetal dat het bestand had bij ontvangst in het CDD+ (of bij het aanmaken, in het geval van een conversie). Is het controlegetal hetzelfde, dan is het bestand ongewijzigd. Zo niet, dan wordt de fout gecorrigeerd door het bestand uit de back-up terug te halen.

Tevens wordt bij elke migratie van de bestanden, zoals het verplaatsen naar een ander medium, de fixity van elk verplaatst bestand gecontroleerd. Optioneel, op verzoek van de archiefvormer, verricht het CDD+ ook een fixity check bij de ontvangst van een bestand, om vast te stellen dat er tijdens het transport geen fouten zijn ontstaan en het verzonden bestand onaangetast is aangekomen. Tot slot kan ook bij het opvragen van een bestand uit het CDD+ een hash worden meegezonden.



(Mogelijke) inzet van fixity checks in e-depot

22

5.3 Bruikbaarheid

De bruikbaarheid van de bestanden wordt geborgd door de conversie naar een duurzaam formaat, zoals hierboven al is beschreven. Deze sectie behandelt een aantal aanvullende punten op het conversiebeleid.

5.3.1 Detectie van bestandsformaat

Computerbestanden hebben allemaal een specifiek bestandsformaat of bestandsindeling. Dit bestandsformaat geeft aan hoe de bits in het bestand geïnterpreteerd moeten worden. Om een bestand goed te interpreteren, is het dus nodig om te weten welk bestandsformaat het heeft. De eerste aanduiding hiervoor

²² De figuur is afkomstig uit CCSDS 650.0-M-3, Reference Model for an Open Archival Information System (OAIS), Magenta Book, december 2024, pagina 4-1.

is de extensie die het gebruikt, zoals .xml voor XML-bestanden of .txt voor een tekstbestand. Deze extensie is echter niet voldoende. Inmiddels bestaan er van veel formaten meerdere versies. Zo heeft bijvoorbeeld PDF drie gangbare versies: 1.4, 1.5 en 1.7. Elk van deze versies maakt gebruik van de extensie .pdf, maar software die geschreven is voor versie 1.4 zal niet goed werken voor bestanden met versie 1.7.

De precieze versie van een bestandsformaat is doorgaans opgenomen in het bestand zelf. Om het te achterhalen, moet het dus geopend worden met software die de verschillende bestandstypes kent. Nadat een bestand is ontvangen, stelt het CDD+ het formaat van dat bestand vast. Aan de hand hiervan kunnen de juiste vervolgacties worden gekozen.

5.3.2 *Validatie en problematische eigenschappen*

Het bestandsformaat stelt allerlei regels aan hoe het bestand moet zijn opgebouwd. Een bestand is valide als het voldoet aan deze regels. Een bestand dat niet voldoet aan deze regels is niet valide. Als een bestand niet valide is, dan kan het niet meer volledig juist worden geïnterpreteerd. Vaak is het bestand nog wel bruikbaar; de software doet dan bepaalde aannames over wat de bedoeling is. Risico hierbij is dat het bestand niet wordt weergegeven zoals het bedoeld is. Als het bestand te ver afwijkt van de standaard, of noodzakelijke gegevens ontbreken, dan is het bestand niet te openen.²³

Van niet-valide bestanden kan de duurzaamheid minder goed worden geborgd. Immers, om de bestanden te interpreteren moeten bepaalde aannames worden gedaan. De software die in de toekomst wordt gebruikt, hanteert mogelijk andere aannames, waardoor niet-valide bestanden anders getoond kunnen worden, of zelfs niet meer geopend kunnen worden. Daarnaast past het CDD+ conversie toe, om bestanden in een niet-duurzaam formaat om te zetten naar een duurzaam formaat. Ook deze conversie loopt tegen problemen aan als bestanden niet-valide zijn, omdat hierbij dezelfde aannames moeten worden gemaakt om de conversie uit te voeren. Conversiesoftware is doorgaans strikter dan de software om bestanden te tonen (dat wil zeggen dat er geen of minder aannames worden gedaan), wat ertoe leidt dat niet-valide bestanden vaak niet geconverteerd kunnen worden.

Naast het niet-valide zijn van een bestand, is het ook mogelijk dat het elementen bevat die bij conversie verloren gaan. Dit is het geval als de elementen zelf conflicteren met de ideeën achter duurzame toegankelijkheid, of als het duurzame formaat waarnaartoe wordt geconverteerd het betreffende element niet ondersteunt²⁴. Een veel voorkomend voorbeeld van dergelijke problematische elementen betreft interactieve elementen waarbij de informatie die wordt getoond verandert als gevolg van de keuzes van de gebruiker.

Ondanks deze mogelijke problemen worden bestanden die niet-valide zijn of andere problematische eigenschappen hebben, in basis wel opgenomen in het CDD+²⁵. Ook in deze gevallen wordt geprobeerd een duurzame versie te creëren. Echter, dit zal

²³ Een bestand dat niet meer geopend kan worden, wordt vaak "corrupt" genoemd. Dit is dus in feite een verregaande vorm van niet valide zijn. Overigens kan het voorkomen dat een bestand dat door bepaalde software niet meer te openen is, toch nog geopend kan worden met andere software (die andere aannames maakt).

²⁴ Idealiter wordt voor de conversie een doelformaat gekozen dat alle elementen ondersteunt, maar dit zal niet altijd haalbaar zijn.

²⁵ Hier kan in overleg met de archiefvormer worden afgeweken.

niet altijd succesvol zijn. In sommige gevallen slaagt de convertor er niet in om een duurzame versie te creëren, en in andere gevallen wordt er wel een duurzame versie gecreëerd, maar is een deel van de informatie of functionaliteit verloren gegaan. De mate waarin dit problematisch is, zal variëren van bestand tot bestand, maar de aanname is dat het in veel gevallen de niet-perfecte duurzame versie nuttig is wanneer het oorspronkelijke bestand niet langer bruikbaar is.

Het is niet altijd mogelijk om zonder handmatige controle na de conversie vast te stellen of er informatie verloren is gegaan. Daarom wordt bij ontvangst van een bestand (voor zover mogelijk) gecontroleerd of het bestand valide is, en of het problematisch eigenschappen heeft. De uitkomst van deze controles wordt vastgelegd in de metadata van het bestand. Daarmee is wordt zichtbaar gemaakt dat er tekortkomingen kunnen zijn bij het geconverteerde bestand.

Een probleem bij het opsporen van niet-valide bestanden is dat bestandformaten niet altijd een waterdichte beschrijving hebben, waardoor er meerdere interpretaties mogelijk zijn. Daarnaast kan validatiesoftware, net als alle andere software, fouten bevatten. Het is dus mogelijk dat verschillende software tot verschillende conclusies komt ten aanzien van hetzelfde bestand. We gaan in deze gevallen altijd uit van de software die binnen het CDD+ is ingezet – waarbij we die natuurlijk bijstellen als er onterechte conclusies zijn getrokken.

Idealiter worden er zo min mogelijk niet-valide bestanden opgenomen in het CDD+, en worden ook problematische eigenschappen vermeden (hoewel dat niet altijd mogelijk zal zijn). In aanvulling op de hier beschreven acties wordt er daarom ook contact opgenomen met archiefvormers als wordt gedetecteerd dat er veel niet-valide of problematische bestanden worden gedeponereerd, om te kijken of dat voorkomen kan worden.

Net zoals van bestandsformaten, is de bruikbaarheid van digitale handtekeningen niet gegarandeerd. Dit wordt verder besproken in paragraaf 5.4.1.

5.3.3 *Conversie van bijlagen en embedded bestanden*²⁶

Het komt voor dat een bestand een bijlage bevat, of dat bestanden zijn ingesloten (embedded). Als dat het geval is dan wordt niet alleen het hoofdbestand behandeld conform dit preservatiebeleid, maar worden ook alle embedded bestanden op deze manier gepreserveerd (hetgeen ook betekent dat bestanden die weer embedded zijn in de embedded bestanden worden gepreserveerd)²⁷. Dat betekent dat ook deze bestanden worden gevalideerd en geconverteerd, indien van toepassing. Hiermee wordt geborgd dat ook de embedded bestanden bruikbaar blijven (voor zover de betreffende bestandsformaten ondersteund worden).

Dit betekent dat het CDD+ preservatiegegevens bijhoudt over zowel het hoofdbestand en de subbestanden, en mogelijk ook de subbestanden los opslaat, in zowel het oorspronkelijke formaat als een geconverteerd formaat. Echter, deze subbestanden zijn niet los opvraagbaar. Gebruikers kunnen alleen het volledige archiefstuk opvragen. Als de verduurzaamde variant wordt opgevraagd, dan zijn er twee mogelijkheden:

²⁶ Voor een uitvoeriger beschrijving van dit soort bestanden en de omgang ermee, zie de Memo Generieke aanpak van embedded bestanden in het CDD+ v1.0, 24 april 2017.

²⁷ Afhankelijk van hoe dit geïmplementeerd wordt, kan dit betekenen dat embedded bestanden meerdere keren in het CDD+ opgeslagen worden: als onderdeel van het hoofdbestand, en als apart bestand.

1. De verduurzaamde variant van het hoofdbestand wordt uitgeleverd, met daarin embedded de duurzame varianten van de subbestanden (of de oorspronkelijke bestanden, als er geen duurzame variant bestaat).
2. Er wordt een archive bestand, zoals een ZIP-bestand, uitgeleverd met daarin de duurzame variant van het hoofdbestand en de subbestanden (of de oorspronkelijke bestanden voor wanneer er geen duurzame variant voor handen is).

De eerste optie zal niet altijd beschikbaar zijn, en hangt af van de mogelijkheden van het formaat waar het hoofdbestand naar geconverteerd is.

5.4 Authenticiteit

5.4.1 *Detectie en validatie digitale handtekeningen*

De authenticiteit van een document kan worden gewaarborgd door het document te voorzien van één of meer handtekeningen. Net zoals van bestandsformaten, is de bruikbaarheid van digitale handtekeningen niet gegarandeerd. De certificaten waarop ze gebaseerd zijn kunnen verlopen, of de validatieservice waarmee ze gecontroleerd worden kan ophouden te bestaan. In een dergelijke situatie kan nog wel worden vastgesteld dat het document ondertekend is, maar niet of deze handtekening valide is. Om ook deze informatie te preserven, valideert het CDD+ de handtekeningen bij ontvangst, en slaat het resultaat van deze validatie op. Deze informatie blijft bewaard in het CDD+, en blijft daarmee dus ook beschikbaar. Als op een later moment de oorspronkelijke handtekening niet meer bruikbaar is, door veroudering of conversie, dan kan aan de hand van deze informatie de oorspronkelijke situatie nog worden gereconstrueerd.

Bijlage: Indeling van bestandsformaten

Categorie	Formaat naam	Gangbare extensies
1	Portable Document Format Archivable	.pdfa
1	Extensible Markup Language	.xml
2	Bitmap	.bmp
2	Graphics Interchange Format	.gif
2	High Efficiency Image Container	.heic
2	HyperText Markup Language	.html
2	Joint Photographic Experts Group	.jpg
2	MIME HTML	.mht
2	Open Document Format for Office Applications	.odt
2	Portable Document Format	.pdf
2	Portable Network Graphics	.png
2	Microsoft Powerpoint for Windows	.ppt
2	Rich Tekst Format	.rtf
2	Tagged Image File Format	.tif, .tiff
2	Text File	.txt
2	Microsoft Open Office XML	.docx
2	Microsoft Office Word document	.doc
3	Advanced Audio Coding	.aac
3	MPEG 1/2 Audio Layer 3	.mp3
3	MPEG-4 Media File	.mp4
3	Microsoft Outlook Item	.msg
3	Excel Binary File Format	.xls
3	Microsoft Excel for Windows	.xlsx