Preservation Storage Criteria, versie 3 (12 december 2018)-NL

Vertaling d.d. 6 december 2022 ten behoeve van NDE-blogs over opslagtechnieken

Doel van dit document

Dit is een lijst van ontwerpcriteria voor digitale bewaring (preservation storage). De criteria zijn bedoeld voor

- Leveranciers van oplossingen voor digitale preservation Storage of
- Gebruikers van een digitale preservation storage oplossing of onderdelen daarvan, variërend van instellingen die net beginnen met preserveren tot instellingen die al een oplossing gebruiken.

Organisaties dienen de criteria aan te passen aan hun lokale omstandigheden. Sommige criteria zullen niet, of op onderdelen van toepassing zijn. Afhankelijk van de rol die een instelling speelt op het gebied van preservering, zullen zij de criteria vanuit een ander perspectief interpreteren. Het criterium "gedocumenteerde toegang" is bijvoorbeeld gedefinieerd als "zorgt voor onveranderlijke logs en/of rapporten die alle toegangen tot het systeem weergeven". Een aanbieder van storagediensten kan dit lezen als zou hij verantwoordelijk zijn voor het verstrekken van logs en rapporten, terwijl de koper dit criterium kan interpreteren als zou hij de logs en rapporten kunnen ontvangen. Elk van de criteria kan worden geïnterpreteerd als "verstrekken" of "ontvangen", afhankelijk van de rol die men vervult met betrekking tot de preservation storage.

Wat is preservation storage?

Preservation storage ondersteunt digitale opslag - "de reeks beheerde activiteiten die nodig zijn om de continue toegang tot digitale objecten te verzekeren voor zolang als nodig" (DPC, 2015). Preservation storage kan gedeeltelijk worden beschouwd binnen de context van het ISO OAIS Reference Model (CCSDS, 2015). In deze context omvat preservation storage dezelfde functies als de OAIS functionaliteit "archiefopslag", evenals de onderdelen van andere OAIS functionaliteiten die nodig zijn om archival information packages (AIP's) op te slaan, in opslag te houden en uit deze opslag op te halen (Zierau & McGovern, 2014). Bijvoorbeeld:

- Delen van "Preservation planning" die betrekking hebben op het monitoren van technologie voor opslag en oplossingen van bitpreservation, migraties van media, veranderingen van overeenkomsten in gelijke informatieobjecten en de organisatie en technologie die nodig is om het beleid voor preservation op opslagniveau te vervullen;
- Delen van "Gegevensbeheer" die de relatie behouden tussen bewaarde gegevens en identificatie van de gegevens in de vorm van metadata;
- Delen van "Administratie" met betrekking tot beleid en normen voor bewaring op het niveau van de opslag;
- delen van "Ingest" die betrekking hebben op de coördinatie van updates van verschillende weergaven van gegevens op het niveau van de opslag.

Naast de context die door het ISO OAIS-referentiemodel wordt geboden, kan preservation storage worden beschouwd in termen van:

- de zich ontwikkelende technologische omgeving en ondersteunende organisatie; en
- de zich ontwikkelende gemeenschappelijke opvattingen over digitale bewaring.

Hoe kunnen de preserveringscriteria worden gebruikt?

De criteria zijn ontwikkeld als een verzameling ontwerpkenmerken die worden overwogen om te komen tot een preservation storage oplossing. Deze kunnen nuttig zijn voor zowel gebruikers als aanbieders van preservation storage oplossingen. Enkele toepassingen van de criteria zijn:

- Het evalueren en vergelijken van preservation storage oplossingen
- Het ontdekken van hiaten in bestaande preservation storage implementaties.
- Het geven van meer gedetailleerde vereisten voor preservation storage
- Als onderdeel van instructiemateriaal over digitale bewaring.
- Het starten van een discussie met IT en andere relevante organisatieonderdelen over preservation storage.
- Het aanzwengelen van discussies binnen huidige en toekomstige gebruikersgroepen van digitale depots over preservation storage.

Waarmee moet nog meer rekening worden gehouden?

Naast deze criteria moet een individuele gebruiker of instelling rekening houden met specifieke criteria voor de eigen organisatie: de eigen eisen, praktijk, beleid, voorschriften, wetgeving en omgevingsfactoren op het gebied van:

- Vertrouwelijkheid en privacy
- Toegang
- Risicobeheer
- Financieel beheer
- Bedrijfscontinuïteit

Zie de Gebruiksgids voor deze criteria voor aanvullende belangrijke overwegingen, waaronder kosten, risico en onafhankelijkheid.

Hoe zijn de criteria tot stand gekomen?

De criteria zijn oorspronkelijk ontwikkeld door Kate Zwaard, Gail Truman, Sibyl Schaefer, Jane Mandelbaum, Nancy McGovern, Steve Knight en Andrea Goethals ter voorbereiding van een workshop tijdens iPRES 2016 genaamd "What is Preservation Storage?" (Goethals et al., 2016). Later hebben Eld Zierau en Cynthia Wu zich aangesloten bij de auteurs om samen te werken aan een verbeterde versie op basis van feedback uit de community. De Criteria bevinden zich momenteel in versie 3, gebaseerd op feedback van deelnemers aan de Designing Storage Architectures-bijeenkomsten van de Library of Congress in 2016 en 2017, PASIG-bijeenkomsten in 2016 en 2017, een iPRES 2017-workshop en via een Google-groep die is opgericht om het onderwerp te bespreken. De laatste stand van zaken leest u op de pagina van de dpstorage-groep op https://groups.google.com/forum/#!forum/dpstorage.

Nr.	Criterium	Categorie	Beschrijving	Gerelateerde criteria en referenties
1.	Integriteitscontrole	Integriteit van de inhoud	Voert verifieerbare controles uit om veranderingen of verlies in of tussen kopieën op te sporen (bv. herberekening van de checksum, fixity checks (controle op onveranderbaarheid) , identificatie van ontbrekende bestanden)	
2.	Onafhankelijke integriteitscontrole	Inhoudelijke integriteit	Ondersteunt fixity checks door andere partijen, bijvoorbeeld de instelling die eigenaar is van de informatieobjecten	
3.	Kostenefficiency	Kostenoverwegingen	Kost in totaal relatief minder dan andere vergelijkbare oplossingen, doordat bij het ontwerp rekening is gehouden met kostenefficiëntie, bijvoorbeeld door resource pooling en -sharing, multi-tenancy (meerdere gebruikers delen dezelfde toepassingen)	
4.	Energie-efficiency	Kostenoverwegingen	Maakt geheel of gedeeltelijk gebruik van energiebesparende principes en technieken. Vereist bijvoorbeeld minder koeling, verbruikt minder stroom, gebruikt minder systemen (rackspace), zoals bij klimaatbewuste computerinitiatieven	
5.	Opslag gewicht	Kostenoverwegingen	Voldoet aan relevante vereisten voor fysiek gewicht zoals gedocumenteerd in SLA. Het gewicht moet bijvoorbeeld lager zijn dan de draagkracht van een specifieke vloer of constructie	
6.	Past zich aan de vereisten aan	Flexibiliteit	Is in staat de opslaginfrastructuur aan te passen aan veranderende lokale eisen, bijvoorbeeld wettelijke vereisten of auditresultaten	
7.	Geografische beperkingen	Flexibiliteit	Maakt de specificatie van de locatie mogelijk, bijvoorbeeld op basis van geografische regionale of geopolitieke kenmerken	
8.	Aanpasbare replicatie	Flexibiliteit	Ondersteunt door de gebruiker gedefinieerde regels voor het verdubbelen van content, bijvoorbeeld minder kopieën van bepaalde content	
9.	Interoperabiliteit	Flexibiliteit	Omvat componenten van opslag die eenvoudig kunnen worden geïntegreerd met andere systemen en toepassingen (d.w.z. plug and play), maakt bijvoorbeeld gebruik van standaard toegangsprotocollen en semantiek zoals Network File System (NFS), SMB, Rest API's	

10.	Open source	Flexibiliteit	Omvat opslagcomponenten die kunnen worden geïntegreerd met open source tools, systemen en diensten in overeenstemming met de voorkeuren van de organisatie	
11.	Vervangbaarheid	Flexibiliteit	Scheidt de opslaglaag van andere systemen in de digitale bewaaromgeving zodat deze onafhankelijk kan worden vernieuwd of vervangen zonder de gehele infrastructuur aan te tasten	
12.	Onderhoudbaarheid	Flexibiliteit	Maakt het mogelijk de opslagfaciliteit te onderhouden en te wijzigen zonder de beschikbaarheid van informatie te verstoren	
13.	Toegangscontrole	Informatiebeveiliging	Biedt op rollen gebaseerde toegangscontroles voor opslaginfrastructuur, bijv. gebruiker, personeel, beheerder, om ervoor te zorgen dat alleen de juiste mensen de juiste toegangsniveaus hebben	
14.	At-rest server-side encryptie met beheerde sleutels	Informatiebeveiliging	Biedt versleuteling, indien nodig, op de opslaglaag, zonder sleutels die klanten hoeven te beheren	
15.	At-rest server-side versleuteling met zelfbeheerde sleutels	Informatiebeveiliging	Biedt encryptie, indien nodig, op de opslaglaag, waarbij klanten de encryptiesleutels beheren	
16.	Integratie van authenticatie	Informatiebeveiliging	Integreert relevante systemen voor authenticatie waarmee interne en externe gebruikers van het systeem zich aanmelden	
17.	Gecodeerde overdracht	Informatiebeveiliging	Gebruikt altijd een passende versleuteling voor de transportlaag bij het verplaatsen van inhoud	
18.	Geografische onafhankelijkheid	Informatiebeveiliging	Bewaart meerdere redundante kopieën op geografisch gescheiden locaties, op voldoende afstand van elkaar, die niet onderhevig zijn aan dezelfde natuurrampen en door mensen veroorzaakte risico's	
19.	Multi-tenancy	Informatiebeveiliging	Ondersteunt afzonderlijke rollen/regels/toegangscontroles voor afzonderlijke agentschappen/afdelingen/colleges/faculteiten enz.	
20.	Organisatorische onafhankelijkheid	Informatiebeveiliging	Beheert kopieën bij verschillende organisaties, waardoor wordt voorkomen dat één organisatie of individu een risico vormt voor alle kopieën van de inhoud	
21.	Permanente verwijdering	Informatiebeveiliging	Ondersteunt de vereiste verwijdering door geautoriseerde gebruikers, in overeenstemming met beleid en regels, waarbij wordt voorkomen dat verwijderde bestanden kunnen worden hersteld	(SNIA, 2017)

22.	Repliceerbaarheid	Informatiebeveiliging	Heeft gedocumenteerde mogelijkheden om binnen redelijke termijnen redundante, gedistribueerde kopieën van inhoud te maken	
23.	Beveiligingsprotocollen	Informatiebeveiliging	Omvat beschermende maatregelen, controles en gedocumenteerde procedures om beveiligingsincidenten te voorkomen met betrekking tot hardware, software, personeel en fysieke structuren, zones en apparaten.	
24.	Rapportage van systeemfouten	Informatiebeveiliging	Biedt onveranderbare logboeken en/of rapporten die alle systeemfouten, storingen en andere kritieke systeemactiviteiten weergeven.	
25.	Technische onafhankelijkheid	Informatiebeveiliging	Bewaart afzonderlijke kopieën in verschillende technische oplossingen (platforms, software inclusief besturingssystemen, hardware, configuraties) om te voorkomen dat alle informatieobjecten worden beschadigd door bijvoorbeeld malware, bugs of andere zwakke punten die met een bepaalde technologie samenhangen.	
26.	Virus/malwaredetectie	Informatiebeveiliging	Omvat software die regelmatig viruscontroles en malwaredetectie uitvoert.	
27.	Herstel bij virussen/malware	Informatiebeveiliging	Biedt herstelacties voor inhoud met virussen en/of malware, bijvoorbeeld quarantaine, kennisgeving, enz.	
28.	Diverse soorten opslagmedia	Veerkracht	Gebruikt verschillende typen opslagmedia/configuraties/providers samen zodat de gewenste niveaus van onafhankelijkheid worden bereikt	(DP Storage WG, 2018, onderdeel Onafhankelijkheid)
29.	Duurzame media	Veerkracht	Biedt gedocumenteerde en aanvaardbare levensduur, storingspercentages en technische kenmerken van de componenten van de opslagmedia	
30.	Controle op fouten	Veerkracht	Voert 24/7/365 foutenopsporing en -correctie uit (bv. met behulp van RAID, Erasure-codering, ZFS, drievoudige kopieën/herbouw)	
31.	Hoge beschikbaarheid	Veerkracht	Heeft een hoog percentage uptime, d.w.z. is operationeel gedurende lange tijd, dankzij technieken zoals het elimineren van single points of failure door redundante apparatuur, load-balanced systemen en effectieve monitoring om software- of hardwarestoringen te detecteren.	(SNIA, 2017)
32.	Hoog aanpassingsvermogen	Veerkracht	Past zich aan onder stress of storingen (bijv. is bestand tegen uitval van apparatuur, stroomuitval, cyberaanvallen, pieken in de vraag van gebruikers)	

33.	Herstel en reparatie	Veerkracht	Herziet, vervangt of repareert ontbrekende of corrupte bestanden binnen aanvaardbare termijnen, zodat geen fouten worden verspreid of biedt de mogelijkheid en de instrumenten om deze acties onafhankelijk uit te voeren, bv. door de eigenaar van de inhoud zelf	Criteria voor kennisgeving van gegevensfouten, transparantiecriteria voor zelfherstel
34.	Volledige export	Schaalbaarheid en prestaties	Ondersteunt de bulkexport van content en metadata om welke reden dan ook, tegen een aanvaardbare snelheid, bijvoorbeeld als onderdeel van een exitstrategie	
35.	Computerkracht	Schaalbaarheid en prestaties	Voldoet aan gespecificeerde rekenkracht voor het systeem of de dienst zoals gedocumenteerd in de SLA	
36.	Levering	Schaalbaarheid en prestaties	Voldoet aan de verwachtingen voor levering vanuit de opslaglaag, bv. tegen een redelijke, onderhandelde snelheid en ter ondersteuning van gelijktijdige gebruikers	
37.	Grenzen van het bestandssysteem	Schaalbaarheid en prestaties	Kan langere bestands-, pad- of mapnamen ondersteunen; omvangrijk aantal bestanden in een map en diverse karaktersets	
38.	I/O-prestaties	Schaalbaarheid en prestaties	Voldoet aan gespecificeerde/onderhandelde input/output prestatieniveaus voor het systeem of de dienst zoals gedocumenteerd in de SLA	
39.	Meerdere opslagniveaus	Schaalbaarheid en prestaties	Ondersteunt het gebruik van meerdere opslagniveaus met verschillende beschikbaarheidsniveaus, bv. on-line, near-line, off-line	
40.	Schaalbaar naar grotere hoeveelheden gegevens	Schaalbaarheid en prestaties	Is in staat om zeer grote hoeveelheden content te ondersteunen qua aantal en omvang van bestanden en totaalvolume	
41.	Ondersteunt uitbreiding	Schaalbaarheid en prestaties	Kan de opslagcapaciteit mettertijd naar behoefte uitbreiden in overeenstemming met eventuele SLA's	
42.	Ondersteunt inperking	Schaalbaarheid en prestaties	Kan indien noodzakelijk de opslagcapaciteit in de loop van de tijd verminderen	
43.	Gelaagde prestaties	Schaalbaarheid en prestaties	Voldoet aan gespecificeerde/onderhandelde prestatieniveaus die passen bij de opgeslagen content, bv. Tier1-opslag voor metadata-indexering en zoeken, Tier2 voor de caching, Tier3 of lager voor bulkopslag.	
44.	Toegankelijkheid	Ondersteuning	Geeft mensen met een beperking gelijkwaardige toegang tot rapporten, documentatie en andere content.	

45.	Onafhankelijke preserveringsdiensten	Ondersteuning	Ondersteunt diensten voor digitale bewaring (bijv. migratie en transformaties met controleerbare resultaten) door andere partijen of externe tools	
46.	Ondersteuning overeenkomst	Ondersteuning	Documenteert de overeenkomst om de opslaginfrastructuur te ondersteunen, bv. door middel van SLA's (waarin verantwoordelijkheden, borging van gegevens, responstijden, bepalingen inzake het beëindigen van de dienst enz. aan de orde komen)	
47.	Opleiding	Ondersteuning	Verstrekt de nodige opleiding aan het betrokken personeel voor alle relevante operationele- en onderhoudstaken.	
48.	Activiteitencontrole	Transparantie	Ondersteunt de mogelijkheid om activiteit in de opslaginfrastructuur te observeren of te controleren (bijv. real-time inzien, logboeken onderzoeken, de performance observeren, de algemene status bepalen of inzoomen op activiteiten)	
49.	Activiteitenrapportage	Transparantie	Biedt rapporten over activiteit in de opslaginfrastructuur (bv. resultaten van fixatie of virussen, corruptie, vervanging door goede kopieën)	
50.	Audits toestaan	Transparantie	Ondersteunt onafhankelijke audits van de storage infrastructuur en -praktijken in overeenstemming met de SLA	
51.	Informatie over beoordelingen	Transparantie	Verstrekt informatie die nodig is ter ondersteuning van beoordelingen, certificeringen, audits en andere bedrijfsactiviteiten door middel van bijvoorbeeld documentatie, rapporten of wegwijzers	
52.	Rapportage over inhoud	Transparantie	Verschaft rapporten over de content in de opslaginfrastructuur (bijv. aantal objecten/bestanden/formaten, gemiddelde bestandsgrootte, soorten objecten, omvang van de gebruikte opslag)	
53.	Rapportage op maat	Transparantie	Ondersteunt aangepaste (bijvoorbeeld configureerbare en/of on-demand) rapportage van content of activiteiten in de opslaginfrastructuur	
54.	Melding van fouten in de data	Transparantie	Stelt eigenaars van content op de hoogte van alle gegevensfouten, herstelacties en problemen binnen redelijke/verwachte/onderhandelde termijnen	
55.	Gedocumenteerde toegang	Transparantie	Biedt onveranderbare logboeken en/of rapporten die alle toegangen tot het systeem weergeven	

56.	Gedocumenteerde infrastructuur	Transparantie	Biedt volledige, complete, actuele en beschikbare documentatie van belangrijke processen, diensten, systemen, procedures, de bekende beperkingen en functies	
57.	Gedocumenteerde herkomst	Transparantie	Documenteert audit/herkomst informatie over alle wijzigingen, bijvoorbeeld over mislukte integriteitscontroles, verwijderingen, wijzigingen, toevoegingen, bewaringsacties en wie of wat de acties heeft uitgevoerd	
58.	Locatie tonen	Transparantie	Toont de specifieke opslaglocatie van gegevens om te voldoen aan SLA-eisen	
59.	Beheer over de niveaus van opslag heen	Transparantie	Ondersteunt beheer en monitoring over meerdere niveaus van beschikbare opslag, bv. on-line, near-line, off-line	
60.	Open opslagformaten	Transparantie	Ondersteunt open-, standaard-, niet aan eigendomsrechten gebonden opslagformaten, bv. TAR, AXF, LTFS	
61.	Zelfherstellende transparantie	Transparantie	Biedt eigenaren van content documentatie over of kennisgeving van elke automatische correctie of wijziging van gegevens om te voldoen aan SLA-eisen.	

Referentielijst

Digital Preservation Coalition. (2015). Digital Preservation Handbook (2nd ed.). Opgehaald van https://www.dpconline.org/handbook

Goethals, A., Knight, S., Mandelbaum, J., Zwaard, K., McGovern, N., Truman, G., & Schaefer, S. (2016) What is Preservation Storage?. Workshop gehouden op de Dertiende Internationale Conferentie over Digitale Bewaring, Bern, Zwitserland. Abstract opgehaald van https://phaidra.univie.ac.at/detail_object/o:502812

DP Storage Working Group (DP Storage WG). (2018). User Guide for the Preservation Storage Criteria (3rd ed.). Opgehaald van https://osf.io/uqkpb

Storage Networking Industry Association. (2017). The SNIA Dictionary (18e ed.) [PDF]. Opgehaald van https://www.snia.org/education/dictionary

The Consultative Committee for Space Data Systems (2012). Reference Model for an Open Archival Information System (OAIS) (CCSDS 650.0-M-2) [PDF] Washington DC: CCSDS. Opgehaald van http://public.ccsds.org/publications/archive/650x0m2.pdf

Zierau, E., McGovern, N. (2014). Supporting Analysis and Audit of Collaborative OAIS's by use of an Outer OAIS - Inner OAIS (OO-IO) Model. [PDF] In Proceedings of the 11th International Conference on Preservation of Digital Objects (iPres) 2014, pp. 209-218. Opgehaald van http://www.ipres-conference.org/ipres14/sites/default/files/upload/iPres-Proceedings-final.pdf