

Eindrapport Marktconsultatie Webarchivering

1. De [Richtlijn archiveren overheidswebsites](#) is de norm voor het archiveren van openbare overheidswebsites
 - a. Kunt u voldoen aan de eisen uit deze norm?

Alle marktpartijen geven aan dat ze voldoen aan de eisen uit de norm voor het archiveren van openbare overheidswebsites. Meerdere marktpartijen noemen de ISO 28500:2017 als standaard voor WARC-files.

- b. Biedt uw oplossing aanvullende functionaliteiten; zo ja welke? En welke toegevoegde waarde heeft/hebben deze functionaliteit(en) voor de zorgdrager?

Marktpartijen benoemen:

- Aanvullende functionaliteiten, bijvoorbeeld de archivering sociale media;
- Beschikbaarheid van een service-portaal waar de zorgdrager een actieve rol kan uitvoeren op het beheer en kwaliteit van de gearchiveerde websites.

Aanvullend worden ook de mogelijkheden tot doorzoekbaarheid van de gearchiveerde websites benoemd en de mogelijkheid tot wetenschappelijk onderzoek op de inhoud.

- c. Staan er eisen in de richtlijn die volgens u niet haalbaar zijn of op praktische bezwaren stuiten? Kunt u dit toelichten?

Bij dit punt worden door meerdere marktpartijen de volgende bezwaren benoemd:

1. Dynamische websites, het gebruik van complexe javascript en embedded content van derde partijen (kaarten, links naar video's);
2. Metadata en de conversie van de ene (oudere) metadata-standaard naar de andere (MDTO). Het is belangrijk om hier voldoende tijd en een validatieproces voor in te richten;
3. Hoe om te gaan met informatiebeveiliging in relatie tot publieke toegang tot de gearchiveerde content. Er is behoefte aan meer toelichting op deze eis;
4. Het ongecomprimeerd opslaan van volledige WARCs: er zou meer aandacht moeten naar duurzame bestandstypes om toegankelijkheid van de inhoud te kunnen waarborgen;
5. De eis van één keer een volledige harvest: volgens één marktpartij zou dit vaker moeten gebeuren om hanteerbare archieven te creëren.

2. Wat zijn volgens u de voor- en nadelen van de opdracht opdelen in meerdere percelen?

Over het algemeen noemen de marktpartijen meer nadelen dan voordelen. Velen noemen als voordeel de betere marktwerking en de mogelijkheid om bij specifieke functionaliteit de beste oplossingen te contracteren. Bij de nadelen worden vooral genoemd de risico's van het beheer van meerdere contracten en de waarborging van het gewenste resultaat (kwaliteit).

3. Wat voor manieren voor afbakening van percelen zijn er volgens u?

De marktpartijen zijn niet eenduidig in de manieren waarop de afbakening van percelen ingericht kan worden. Het varieert van:

- Indeling van percelen op basis van functionaliteit (bijvoorbeeld: Harvesting – Opslag – Toegang – Migratie en Transitie);
- Indeling van percelen op basis van organisatie / zorgdrager (bijvoorbeeld: clusters van ministeries en ZBO's);
- Indeling van percelen op basis van complexiteit van te archiveren websites.

4. Gaat uw voorkeur uit naar één perceel of naar meerdere percelen? Waarom?

De meeste marktpartijen geven de voorkeur aan om de opdracht in één perceel aan te bieden vanwege beheersbaarheid en uniforme kwaliteit en toegang. Er zijn ook marktpartijen die de wens uitspreken om het wel in percelen op te knippen vanwege de benodigde expertise. Ook zijn er partijen die flexibel zijn en kunnen acteren op beide scenario's (opdeling opdracht in één of meerdere percelen).

5. Is uw organisatie ISO gecertificeerd? Zo ja, welke certificering(en)?

Alle marktpartijen op één na zijn ISO-gecertificeerd.

Genoemde certificeringen en gekoppelde thema's:

- NEN-ISO/IEC 27002 (cybersecurity, operationele security)
- ISO/IEC 27001 (softwareontwikkeling, infrastructuurbeheer, organisatie, archivering, support)
- ISO 27001, SOC 2 Type II, and C5 certificaties (dataopslag)
- ISO 9001 (wordt door één bedrijf genoemd)
- ISO 28500:2017 (WARC standaard)
- BIO (Baseline Informatiebeveiliging Overheid)

6. Datasoevereiniteit is bij Webarchivering een belangrijk thema. Met het oog door ons gewenste datasoevereiniteit willen we Webarchivering alleen toestaan op servers waarvan housing en hosting plaatsvindt in de Europese Economische Ruimte.

a. Kunt u aan dit verzoek voldoen?

Alle marktpartijen op één na kunnen aan het verzoek voldoen.

b. Zo nee: waarom niet?

Een marktpartij maakt een onderscheid tussen gegevenssoevereiniteit en gegevensopslag en hanteert een datacentrische visie op soevereiniteit waarbij de wetgeving de data volgt; ook al wordt de data in andere regio's verwerkt.

7. Op grond van welke factoren (bijvoorbeeld: harvest, seed, crawl, opslag, ...) bepaalt u de kosten voor een te archiveren website?

De marktpartijen geven uiteenlopende antwoorden, de één wat uitgebreider dan de ander. Over het algemeen worden componenten als aantal en grootte van websites, complexiteit van de te archiveren website, de opslagcapaciteit en kwaliteitsbeheersing genoemd. Slechts een van marktpartijen ziet het gebruik (viewer usage) van het archief als factor.

8. Kunt toelichten welke kostencomponenten vast en variabel zijn?

De kostenstructuur van de verschillende marktpartijen lopen uiteen. Bij sommige marktpartijen liggen alle kosten vast (all-in prijs, vaak met een beperking op aantal/volume), bij andere marktpartijen is alles variabel.

- Meest genoemde vaste kosten: inrichting van service-portaal, cybersecurity en compliance, licentie, Service Level Agreement (SLA) borging, overdracht naar Nationaal Archief (voor een afgesproken tarief), hosting van viewer. Harvest van statische websites is meestal ook vast.

- Meest genoemde variabele kosten: aantal websites, complexiteit (gemiddeld, complex), benodigde rekenkracht / tijdsduur, opslagvolume, nabewerking, import bestaande webarchieven plus eventuele bewerking, plus eventuele bewerking (reharvest), ondersteuning / handmatige bewerking / kwaliteitszorg, metadatamanagement.

9. Hoe wordt de automatische kwaliteitscontrole, op kwaliteit van de gearchiveerde website en op de technische kwaliteit van WARC-bestanden, ingeregeld?

Antwoorden lopen uiteen van uitgebreid tot summier tot geen nadere informatie vanwege het publieke karakter van de marktconsultatie. De meeste marktpartijen maken geautomatiseerde rapportages op basis van de archiveringsketen waarbij enkele marktpartijen de uitkomsten van de rapportages ook inzichtelijk maken voor de opdrachtgever/zorgdrager via het service-portaal.

10. Kunt u beschrijven hoe de webarchiefbestanden en metadata tijdens en na de archivering op juistheid en volledigheid worden gecontroleerd?

a. Wat zijn de specifieke aandachtspunten waarop wordt gecontroleerd?

Veel marktpartijen voeren dit uit op basis van de standaard (ISO 28500:2017). Een partij geeft aan dat er ook periodiek wordt gecontroleerd of oudere archieven technisch en functioneel correct blijven werken. Daarnaast is er een partij die periodiek de integriteit controleert.

b. Licht toe hoe u [ISO 28500:2017](#) toepast in de dienstverlening

Marktpartijen geven met name aan dat ze deze standaard toepassen. Een tweetal marktpartijen neemt actief deel aan het IPCC (International Internet Preservation Consortium). Hoe deze standaard wordt toegepast wordt beschreven in relatie tot andere technische vereisten waaraan moet worden voldaan.

c. Kunt u beschrijven hoe de resultaten van deze controle worden gecommuniceerd met de zorgdragers?

Marktpartijen reageren verschillend in frequentie, inhoud en manier waarop er gecommuniceerd wordt. Veel partijen bieden communicatie aan via een service-portaal. Er is er ook een marktpartij waar een service-portaal tot de mogelijkheden behoort en waar dat service-portaal in ontwikkeling is.

11. Op welke wijze rapporteert u over:

a. Mogelijk incorrecte (bijvoorbeeld: fouten in URL, lege websites of opdrachten die leiden tot lege WARC-bestanden et cetera) Webarchiveringsopdrachten?

Bijna alle partijen geven aan deze zaken automatisch te loggen. Rapportage gebeurt deels in een service-portaal (zie ook het antwoord op vraag 1b) en deels (soms aanvullend) per mail (incidenteel opgevolgd door een telefoontje).

b. Duplicaten (=twee zorgdragers die dezelfde URL aanleveren)?

Het rapporteren over duplicaten is meestal standaard, maar moet soms wel worden geactiveerd. Veel partijen hebben een centraal URL-register dat toevoegen van een duplicaat verhindert, c.q. daarover beide zorgdragers direct informeert. Een partij zegt dat het opvolgen van de Richtlijn duplicaten onmogelijk maakt.

c. Niet harvestbare content: onvolkomenheden bij (de archivering van) websites zoals missende onderdelen?

Een marktpartij zegt dat deze content geïdentificeerd kan worden met kwaliteitschecks. Een tweede dat deze op het service-portaal kan worden ingezien. Een derde dat deze tijdens de initiële harvest en daarna wordt gedetecteerd en daarover afstemming plaatsvindt met de zorgdrager. En een vierde stelt dat informatie hierover ook opgeslagen wordt in een ODS-spreadsheet. Ook wordt gewezen op het kwaliteitskader voor metadata en archiefbestanden dat gebruikt wordt.

d. Harvests die buiten de scope gaan, zoals bijvoorbeeld: crawler traps, eeuwigdurende kalenders, boodschappenmandjes, et cetera?

Veel marktpartijen geven alleen aan hoe ze deze zaken (technisch) oplossen. Een enkeling geeft aan dat rapportage plaatsvindt via het service-portaal of dat structurele afwijkingen worden gerapporteerd (zonder duiding op welke wijze) en besproken.

12. De richtlijn archiveren overheidswebsites schrijft voor welke metadata moeten worden meegeleverd met de archiefbestanden tijdens overbrenging. De meeste metadata moeten worden geleverd door de marktpartij. Er zijn ook metadata die geleverd moeten worden vanuit de archiefvormende organisatie (denk bijv. aan een beschrijving van de website en de informatiecategorie van de selectielijst waar het websitearchief onder valt). Deze metadata moeten worden samengebracht in één export bij overdracht van het archief.

a. Heeft uw organisatie ervaring met het genereren van deze metadata? Kunt u aangeven welke?

Op één na hebben alle marktpartijen hiermee ervaring. De partij die dit niet heeft, geeft aan dat zij een digital preservation of asset management systeem bij de zorgdrager veronderstelt ten behoeve van het beheer van metadata. Een aantal marktpartijen geeft aan ook metadata te beheren, naast die te kunnen genereren. Eén partij vermeldt daarbij expliciet het samenbrengen van metadata die de archiefvormende organisatie levert met de automatisch gegenereerde metadata uit het archiveringsproces, ten behoeve van de export van het archief.

b. Welke mogelijkheden biedt u om metagegevens te controleren en toe te voegen?

Eén partij geeft aan op verschillende manieren en in verschillende fasen controle mogelijkheden te bieden:

- een service-portaal met validatieondersteuning voor de zorgdrager;
- een *validation engine* die aan de hand van het relevante metadata-profiel toetst, voorafgaand aan export;
- een compliance rapport dat voorafgaand aan een export wordt gegenereerd.

Andere partijen bieden de mogelijkheid van een service-portaal om metadata in te zien, voor validatie en verrijking. Beheer per periode wordt door één partij expliciet benoemd.

Twee partijen bieden geen mogelijkheden om metadata te controleren of toe te voegen.

c. Deze metadata zal MDTO-conform moeten zijn. Kunt u toelichten hoe u hieraan kunt voldoen?

Alle marktpartijen zeggen hieraan te kunnen voldoen, in redelijk algemene termen, maar niet in detail hoe zij dat daadwerkelijk zullen bewerkstelligen. Twee partijen noemen voorzieningen voor de overgangsperiode van ToPX naar MDTO. Eén partij geeft aan dat tijdens deze periode het systeem zowel MDTO als ToPX zal genereren en is bereid dit te demonstreren. Een partij zegt een e-depot met MDTO model te hebben.

d. Heeft uw organisatie ervaring met het overdragen van deze data naar een archiefbewaarplaats c.q. aan een andere dienstverlener? Kunt u aangeven welke?

Alle marktpartijen zeggen deze ervaring te hebben. Enkel maken dit concreet, onder meer door het noemen van een integriteitscheck. Eén partij verwijst naar het gebruik van een SIP, een ander naar vloeiend overbrengen. Eén partij kan ontvangen en overdragen, maar heeft met dat laatste geen ervaring.

13. In sommige gevallen moet een deel van een gearchiveerde website worden uitgesloten van beschikbaarstelling in een viewer vanwege privacy, auteursrecht of informatiebeveiliging.

a. Hoe kan gearchiveerde content beperkt beschikbaar worden gemaakt? Kan een zorgdrager dat ook zelf?

Alle marktpartijen bieden deze mogelijkheid en de meeste geven aan hoe die eruitzien. Veel partijen geven aan dat zorgdragers dit (ook) zelf kunnen regelen in het service-portaal van de marktpartij. Het detailniveau van de beperking verschilt. Sommigen spreken over specifieke URL's en hele websites, anderen spreken over granulaire mogelijkheden of uitsluiting van beschikbaarstelling op basis van aanpassing van de metadata van het document.

b. Hoe wordt die beperkte beschikbaarheid (conform 4.4 in de [Richtlijn archiveren overheidswebsites](#)) geadministreerd?

Twee partijen verwijzen naar de richtlijn en de daarin opgenomen ODF-spreadsheet voor administratie. Een aantal partijen geeft aan dat dit via het service-portaal (via role-based access voor geautoriseerde gebruikers) mogelijk is.

c. Op welke wijze vindt overdracht van onder @b genoemde administratie plaats naar een mogelijke volgende dienstverlener c.q. de archiefbewaarplaats?

Alle partijen geven aan dat het exporteren met metadata van @b mogelijk is. Het merendeel geeft aan dat dit via een exportbestand zoals .csv of soortgelijk bestand gaat. Een paar partijen noemen niet het bestandstype voor de export.

14. De aanbesteding omvat mogelijk ook het in ontvangst nemen en het aansluitende beheer van en toegang tot data van eerder gecreëerde webarchieven.

a. Bent u in staat om deze eerder gecreëerde webarchieven in ontvangst te nemen en te verwerken?

Welke eisen en beperkingen (technisch, metadata, hoeveelheid data) stelt u aan het te over te dragen archief?

Alle marktpartijen geven aan eerder opgebouwde webarchieven in ontvangst te kunnen nemen. Vaak wordt als eis genoemd dat de WARC files ISO 28500:2017 compliant moeten zijn en/of wordt gerefereerd naar de WARC 1.1 standaard en/of wordt verwezen naar eisen in de Richtlijn archiveren overheidswebsites. Een enkele partij wijst op beschikbaarheid van metadata conform deze richtlijn. Een andere voorwaarde die enkele keren wordt genoemd is een veilige overdrachtsmethode, zoals bijvoorbeeld: sFTP.

b. Heeft u ervaring met het ontvangen, indexeren van en toegang bieden tot dergelijke hoeveelheden data (>1 Petabyte)? Kunt u daar voorbeelden van geven?

Vrijwel alle marktpartijen hebben hiermee ervaring, soms op een wat kleinere (dan 1 Petabyte) schaal. Een enkele partij noemt als bron andere marktpartijen die webarchieven maken. Op twee partijen na geven marktpartijen geen concrete voorbeelden. Een partij geeft aan dat die voorbeelden wel gegeven kunnen worden als er een NDA wordt overeengekomen.

- c. Kunt u de door u gevormde webarchieven (en indien van toepassing de eerder gecreëerde webarchieven) overbrengen naar het Nationaal Archief, met inachtneming van de [voorwaarden](#) die het Nationaal Archief stelt aan het overbrengen van digitale archieven; in het bijzonder aan het [Voorwaarden export document](#).**

Alle partijen zeggen dat ze dit kunnen. Een partij adviseert continue overdracht van archieven gedurende de looptijd van het contract.

15. Staat u open voor een gesprek over deze Marktconsultatie met de Aanbestedende Dienst?

Alle partijen staan open voor een gesprek.