

Richtlijn archiveren overheidswebsites voor publiekscommunicatie

Versie 0.3

Datum 14 mei 2018
Status Concept

1

2

CONCEPT

3 Colofon

4 Contact F +31-70-331 5460
5 contact@nationaalarchief.nl
6 Postbus 90520
7 2509 LM Den Haag
8 E-Docs nr #1265833

9

| Versiegeschiedenis | | |
|---------------------------|--------------|---|
| <i>Versie</i> | <i>Datum</i> | <i>Beschrijving</i> |
| 0.1 | 06-12-2017 | Eerste versie. Als concept gepubliceerd op KIA en voor een eerste consultatie besproken met de direct betrokkenen bij het opstellen en de eerste implementaties van de richtlijn. |
| 0.2 | 07-05-2018 | Tweede versie waarin eerste opmerkingen van direct betrokkenen verwerkt zijn. Ter informatie voorgelegd aan CIO-Rijk, SG OCW en de Algemeen Rijksarchivaris. |
| 0.3 | 14-05-2018 | Publicatie reviewversie op KIA en start openbare review. |

10

CONCEPT

| | | |
|----|---------------|---|
| 11 | Inhoud | |
| 12 | 1 | Over deze richtlijn 4 |
| 13 | 2 | Inleiding..... 5 |
| 14 | 3 | Doel, doelgroepen en scope 6 |
| 15 | 4 | Waarom webarchivering? 8 |
| 16 | 5 | Eisen aan webarchivering 9 |
| 17 | 6 | Bijlage: Voorbeeld stappenplan invoeren webarchivering.....17 |
| 18 | 7 | Bijlage: Begrippen23 |
| 19 | 8 | Bijlage: Motivatie van gemaakte keuzes32 |
| 20 | 9 | Bijlage: Aanleiding en opdracht van deze richtlijn35 |
| 21 | | |

CONCEPT

1 22 Over deze richtlijn

23 Deze richtlijn helpt overheidsorganisaties bij het archiveren van hun websites voor
24 publiekscommunicatie. Zo kunnen ze voldoen aan de Archiefwet. De richtlijn
25 beschrijft inzichten en werkwijzen die zich in de praktijk hebben bewezen. Het
26 toepassen ervan is vrijwillig, overheidsorganisaties hoeven deze richtlijn niet te
27 volgen. Maar dan zullen ze wel op een andere manier moeten voldoen aan de
28 verplichtingen uit de Archiefwet.

29 Het Nationaal Archief schreef deze richtlijn, in opdracht van de minister van OCW.
30 Bij het schrijven werkten we nauw samen met veel betrokken partijen en experts.
31 Er volgt nog een openbare review op deze conceptversie. Zo kunnen alle
32 geïnteresseerden onduidelijkheden aangeven en verbeteringen voorstellen.

33 De Standaardisatieraad van het Nationaal Archief ziet erop toe dat het Nationaal
34 Archief de richtlijn zorgvuldig verder ontwikkelt. Het resultaat moet voor alle
35 belanghebbenden inhoudelijk acceptabel en praktisch bruikbaar zijn. In de
36 Standaardisatieraad zijn alle lagen van de overheid en verschillende
37 archiefinstellingen vertegenwoordigd.

38 Het archiveren van websites is nog geen standaardpraktijk. Het toepassen van deze
39 richtlijn zal daarom in de praktijk nog vragen oproepen en problemen opleveren.
40 Met die praktijkervaringen kunnen we de richtlijn verbeteren en uitbreiden.
41 Organisaties die een casus uit de praktijk willen inbrengen, kunnen daarvoor een
42 mail sturen naar contact@nationaalarchief.nl.

CONCEPT

2 43 Inleiding

44 Het is voor overheidsorganisaties belangrijk om hun websites te archiveren. Zo blijft
45 informatie bewaard die later onmisbaar kan zijn. Bovendien verplicht de Archiefwet
46 het. Het maakt daarbij niet uit van wie het platform is waarop de website beheerd
47 wordt. Een overheidsorganisatie die verantwoordelijk is voor de inhoud van een
48 website, is ook verantwoordelijk voor het archiveren van die inhoud.

49 Maar hoe archiveer je een website? Waar begin je en welke onderdelen bewaar je?
50 Wie moet dat doen en hoe vaak? Deze richtlijn geeft antwoorden op die vragen.

51 Omdat websites steeds kunnen veranderen is webarchivering ook een doorgaand
52 proces. Dat zo snel als mogelijk geregeld moet zijn nadat een website actief wordt.
53 Het is onvoldoende om pas maatregelen te nemen op het moment dat een
54 organisatie een website niet meer bijhoudt of vervangt. Want dan zijn wijzigingen
55 uit het verleden moeilijk of niet meer te achterhalen. De richtlijn is daarom van
56 toepassing op alle nieuwe en bestaande websites en schrijft voor dat webarchieven
57 dagelijks worden bijgewerkt.

58 Met webarchivering bedoelen we alles wat nodig is om de inhoud van alle
59 overheidswebsites duurzaam toegankelijk te houden. Dat is te veel om allemaal in
60 deze eerste versie van de richtlijn te behandelen. Daarom richt deze eerste versie
61 zich op harvesting van websites voor publiekscommunicatie. Later breiden we de
62 richtlijn stap voor stap uit.

63 De bestuurder van de overheidsorganisatie waar de website onder valt is
64 verantwoordelijk voor het archiveren van de websites. De bestuurder (zorgdrager)
65 ziet erop toe dat de archivering gebeurt, ook als hij deze taak of delen ervan
66 uitbesteedt aan andere partijen.

CONCEPT

3 67 Doel, doelgroepen en scope

3.1 68 Doel

69 Het doel van deze richtlijn is dat overheidsorganisaties de kennis hebben om hun
70 websites te archiveren en deze kennis toepassen. Daardoor blijft de inhoud van
71 overheidswebsites toegankelijk voor iedereen die daar belang bij heeft. En zo
72 kunnen overheidsorganisaties zich houden aan de Archiefwet.

3.2 73 Doelgroepen

74 Deze richtlijn is van toepassing op alle overheidsorganen zoals bedoeld in de
75 Archiefwet. Zowel binnen de centrale als decentrale overheden.

76 Met overheidsorganen worden ook bedoeld particuliere instellingen die wettelijke
77 taken uitvoeren. Maar de richtlijn geldt alleen voor die wettelijke taken, bijvoorbeeld
78 het verstrekken van diploma's door opleidingen en Apk-keuringen door garages.

79 Binnen deze overheidsorganen zijn er twee groepen mensen waar deze richtlijn zich
80 speciaal op richt:

- 81 1. Opdrachtgevers van websites.
- 82 2. De mensen die de opdrachten vertalen in plannen en die (laten) uitvoeren.

83 De opdrachtgevers zijn verantwoordelijk voor de inhoud van de website, en
84 daarmee ook voor het archiveren daarvan. Met deze richtlijn kunnen zij opdrachten
85 geven voor het archiveren en toezien op de uitvoering.

86 Deze richtlijn is ook bedoeld voor de mensen die de opdrachten vertalen in plannen
87 en die (laten) uitvoeren. Bijvoorbeeld informatiemanagers en -analisten, functioneel
88 beheerders, architecten en (inkoop)adviseurs. Deze richtlijn helpt hen bij het
89 uitwerken van de opdracht om websites te archiveren.

3.3 90 Scope

91 Deze richtlijn is bedoeld voor het archiveren van webpagina's voor
92 publiekscommunicatie. Dat zijn webpagina's die openbaar en voor iedereen gelijk
93 zijn. Ook sociale media vallen hieronder, voor zover ze hieraan voldoen.

94 De richtlijn is bedoeld voor het archiveren van eigen websites door
95 overheidsorganisaties. Niet voor het archiveren van websites van anderen door
96 archief- en erfgoedinstellingen.

97 Deze richtlijn gaat uit van de bewezen en betaalbare techniek van harvesting. Dat is
98 het binnenhalen en bewaren van webpagina's zoals de gebruiker ze krijgt
99 aangeleverd. Harvesting heeft een aantal beperkingen waardoor bepaalde
100 webpagina's niet goed gearchiveerd worden (zie paragraaf [7.23](#)). Als daardoor
101 belangrijke informatie verloren gaat, dan zijn aanvullende maatregelen nodig die
102 buiten de scope van deze richtlijn vallen.

103 Informatie over de totstandkoming van webpagina's valt niet onder deze richtlijn.
104 Tenzij deze informatie ook op een webpagina staat. Denk hierbij aan het
105 achterliggende communicatiebeleid, informatie over de techniek waarmee de
106 website is gemaakt en de inhoud van de databases waaruit de webpagina's gemaakt
107 zijn. Ook deze informatie valt onder de Archiefwet en zal mogelijk gearchiveerd
108 moeten worden. Maar daar worden andere technieken voor gebruikt.

CONCEPT

109 Deze richtlijn gaat alleen over webpagina's. Niet over de browsers waarmee
110 bezoekers websites bekijken. Omdat die browsers niet van de overheid zijn, vallen
111 ze ook niet onder de Archiefwet.

CONCEPT

4 112 Waarom webarchivering?

113 Overheden gebruiken websites om informatie te geven aan burgers, bedrijven,
114 instellingen en andere overheden. Die informatie kan in de loop van de tijd
115 veranderen of van de site gehaald worden. Voor verschillende partijen kan het
116 belangrijk zijn dat de informatie dan nog steeds toegankelijk is. Die belangen zijn
117 heel verschillend:

- 118 • *Uitvoeren van bedrijfsprocessen*
119 Bij het uitvoeren van taken kan het nodig zijn om oude webpagina's te bekijken.
120 Bijvoorbeeld voor het beantwoorden van vragen van burgers of het opstellen van
121 rapportages. Ook kan de inhoud van oude websites soms hergebruikt worden.
- 122 • *Verantwoording over overheidshandelen*
123 Websites zijn uitingen van overheidshandelen. Overheidsorganisaties moeten
124 zich daarover kunnen verantwoorden. Daarvoor moet de informatie beschikbaar
125 zijn zoals het ooit openbaar was. Dat kan nodig zijn voor het behandelen van
126 Wob-verzoeken of in een debat. Maar ook bij een parlementair- of
127 raadsonderzoek, een audit of rekenkameronderzoek.
- 128 • *Recht- en bewijszoekende burgers*
129 Burgers en bedrijven kunnen rechten ontleen aan bepaalde informatie op
130 websites. Bijvoorbeeld bij het aanvragen van diensten, vergunningen, subsidies,
131 of het starten van een bedrijf. Burgers en bedrijven baseren keuzes en
132 handelingen op de informatie van de overheid. Die informatie kan daarom nodig
133 zijn als bewijs in beroeps- en bezwaarprocedures.
- 134 • *Onderzoek*
135 Overheidswebsites geven onderzoekers inzicht in hoe overheden hun taken
136 uitvoerden en zichzelf presenteerden. Ze zijn dus ook cultureel erfgoed en een
137 bron voor geschiedschrijving.

138 Daarom archiveren overheidsorganisaties hun websites. Welke websites moeten ze
139 archiveren en voor hoe lang? Dat hangt af van wat voor soort website het is, de
140 belangen die ermee gemoeid zijn en de kosten. Het maken van deze afwegingen is
141 onderdeel van het archiveringsproces.

5 142 Eisen aan webarchivering

143 In deze richtlijn stelt het Nationaal Archief op 12 punten eisen aan goede
144 webarchivering. Ze gaan over alles waar een overheidsorganisatie op moet toezien
145 voor, tijdens en na het archiveren van een website:

- 146 1. Overzicht websites
- 147 2. Beperkt openbaar
- 148 3. Dagelijks harvesten
- 149 4. Volledig harvesten
- 150 5. Bestanden
- 151 6. Metagegevens
- 152 7. Kwaliteitscontrole
- 153 8. Beveiliging
- 154 9. Eigenaarschap
- 155 10. Toegang
- 156 11. Blijvend bewaren
- 157 12. Overbrenging

158 De eisen zijn bedoeld voor individuele overheidsorganisaties. Dat wil niet zeggen dat
159 die alles zelf moeten doen. Een aantal eisen gaan bijvoorbeeld over het
160 harvestingsproces en -resultaat. Wanneer een harvestingsdienstverlener de harvestings
161 uitvoert, hebben overheidsorganisaties bij deze eisen de taak van opdrachtgever. De
162 harvestingsdienstverlener moet zich bij het uitvoeren van de taak houden aan de
163 eisen.

5.1 164 **Overzicht websites**

165 *Eis:* Er is een actueel overzicht met beschrijvingen van alle websites waarvan de
166 overheidsorganisatie verantwoordelijk is voor de inhoud. Als de organisatie
167 verantwoordelijk is voor slechts een deel van een website, dan staat dat deel in het
168 overzicht.

169 In het actuele overzicht met beschrijvingen van websites staat ook welke varianten
170 van de website bestaan. Varianten kunnen als aparte website benaderbaar zijn,
171 bijvoorbeeld door een specifiek stukje in de hostname of de URL van de website. Als
172 dat zo is, dan worden deze als afzonderlijke websites geharvest.

173 Als er afhankelijk van de context van de gebruiker verschillende varianten van de
174 webpagina's worden geleverd, dan zijn deze varianten allemaal opgenomen in de
175 beschrijving. Bijvoorbeeld als er andere pagina's geleverd worden voor verschillende
176 talen of typen apparaten. Varianten van sites voor verschillende talen kunnen
177 bijvoorbeeld gekozen worden op basis van de inhoud van een cookie.

178 Per (deel)website zijn verschillende zaken beschreven:

179 **Algemeen**

- 180 • Naam van de website.
- 181 • Talen van de website.
- 182 • Korte beschrijving van de inhoud (bijvoorbeeld uit "over deze site").

183 **Contenteigenaar**

- 184 • De contenteigenaar (organisatie of organisatieonderdeel) die verantwoordelijk is
185 voor de inhoud van de website.

CONCEPT

- 186 • Link naar de beschrijving van de organisatie in het Actorenregister, als deze
187 daarin staat.

188 **Beheergegevens**

- 189 • Contentbeheerder.
190 • Platformbeheerder.
191 • Platform dat wordt gebruikt voor het beheer.
192 • Datum waarop de website online ging.
193 • Datum waarop website offline ging. Dit geldt alleen voor een site die is gestopt.

194 **Parameters voor harvesting**

195 Per variant van de website:

- 196 • Korte beschrijving van deze variant (bijvoorbeeld "voor mobiele apparaten").
197 • De harvestingscope die aangeeft welke webpagina's en bijlagen tot (deze
198 variant van) de website gerekend worden. Bestaande uit:
199 o Site-URL.
200 o Secundaire site-URL's (indien van toepassing).
201 o De startpagina (seed-URL).
202 o Lijst met weespagina's.
203 o Blacklist, met per webpagina's op de blacklist:
204 ▪ Korte beschrijving van de inhoud van de webpagina.
205 ▪ Toelichting waarom de webpagina niet geharvest kan worden.
206 • Eventuele parameters die meegegeven worden door de harvester voor server-
207 driven content negotiation. Specifiek voor deze versie van de website. Zoals de
208 gewenste taal of het type apparaat waarvoor de webpagina geschikt moet zijn.

209 *Toelichting:* Het overzicht zorgt ervoor dat geen websites vergeten worden bij het
210 archiveren. Bovendien bevat de beschrijving van de sites informatie om ze te
211 kunnen archiveren. En om de gearchiveerde websites later te kunnen vinden.

5.2 212 **Beperkt openbaar**

213 *Eis:* Er is per website een actuele lijst van URL's van webpagina's en bijlagen die
214 tijdelijk beperkt openbaar gemaakt moeten worden. Per item staat op deze lijst:

- 215 • De URL van de website of bijlage waarvoor de beperking geldt.
216 • Het deel van de webpagina of bijlage waarvoor de beperking geldt. Bijvoorbeeld
217 'opsporingsberichten' of 'geheel'.
218 • De wettelijke grondslag waarom de beperking geldt voor overbrenging.
219 Bijvoorbeeld de Wet openbaarheid van bestuur (Wob) en de Algemene
220 verordening gegevensbescherming (Avg).
221 • Archiefwettelijke grondslag waarom de beperking moet gelden na overbrenging
222 (zie brochure 'Open tenzij...'):
223 o Bescherming van de persoonlijke levenssfeer.
224 o Belang van de Staat of zijn bondgenoten.
225 o Onevenredige bevoordeling of benadeling van betrokken natuurlijke
226 personen, rechtspersonen of derden.
227 • Jaartal tot wanneer de beperking geldt.

228 *Toelichting:* Deze richtlijn gaat over openbare websites. Toch kan het voorkomen
229 dat informatie op een website achteraf gezien niet (meer) openbaar mag zijn.
230 Bijvoorbeeld een opsporingsbericht voor een misdrijf of gevoelige informatie die per
231 ongeluk is geplaatst. Er gelden regels voor het beperken van de toegang tot deze
232 informatie, ook in archieven. Deze regels dienen gerespecteerd te worden. In het

CONCEPT

- 233 belang van de betrokken personen en organisaties, en soms de maatschappij als
 234 geheel. De lijst met beperkt openbare website en bijlagen maakt dit mogelijk.
- 235 Webpagina's en bijlagen waarvoor beperkte openbaarheid geldt worden niet
 236 vernietigd. Ze blijven onderdeel van het archief. Alleen de toegang wordt beperkt.

5.3 237 **Dagelijks harvesten**

- 238 *Eis:* Alle websites uit het overzicht worden dagelijks geharvest. Een keer per jaar
 239 wordt er een volledige harvest gemaakt. Daartussen vindt dagelijkse incrementele
 240 harvesting plaats van alleen de wijzigingen.
- 241 *Toelichting:* Harvesting vormt de basis voor de webarchivering. Het legt een website
 242 vast zodat iemand die later nog kan raadplegen. Ook nadat die veranderd is.
 243 Dagelijks harvesten zorgt ervoor dat de meeste wijzigingen gearchiveerd worden.
 244 De jaarlijkse volledige harvesting geeft extra zekerheid voor het geval een eerder
 245 archief verloren is gegaan.

5.4 246 **Volledig harvesten**

- 247 *Eis:* Per website worden alle webpagina's geharvest, met alle componenten en
 248 bijlagen die binnen de harvestingscope vallen. Als er meer varianten zijn van een
 249 website, dan worden die allemaal helemaal geharvest.
- 250 Dat wil zeggen de startpagina en de weespagina's, alle webpagina's en bijlagen die
 251 van daaruit met hyperlinks bereikbaar zijn en alle componenten van de webpagina's
 252 (zoals afbeeldingen en stylesheets). Voor zover het webadres begint met de site-
 253 URL of een van de secundaire site-URL's, en niet op de blacklist staat.
- 254 Bij het doorlopen van de hyperlinks maken het aantal hops en de paddiepte niet uit.
- 255 *Toelichting:* Alle webpagina's en bijlagen dragen bij aan de betekenis van een
 256 website. Daarom worden ze allemaal geharvest. Behalve pagina's op de blacklist
 257 omdat het technisch of inhoudelijk niet mogelijk is deze te harvesten.

5.5 258 **Bestanden**

- 259 Leg de resultaten van harvesting vast in webarchiefbestanden die voldoen aan de
 260 volgende eisen.

5.5.1261 *WARC-standaard*

- 262 *Eis:* Elk webarchiefbestand voldoet aan de standaard WARC NEN-ISO 28500:2017.
- 263 *Toelichting:* Het WARC-formaat is de internationale standaard voor webarchivering.
 264 Door deze standaard te gebruiken is duidelijk wat de inhoud van een
 265 webarchiefbestand betekent en zijn er hulpmiddelen om de bestanden te verwerken.

5.5.2266 *Grootte*

- 267 *Eis:* Elk webarchiefbestand is maximaal 1GB groot.
- 268 Wanneer het totale webarchief groter is dan 1GB moet het worden opgesplitst in
 269 webarchiefbestanden die elk maximaal 1GB groot zijn. Uit de naamgeving van de
 270 bestanden moet blijken welke webarchiefbestanden bij elkaar één webarchief
 271 vormen. Alleen wanneer een individueel WARC-bestand van zichzelf al groter is dan
 272 1GB mag het in een eigen webarchiefbestand zijn opgeslagen dat groter is dan 1GB.

CONCEPT

273 *Toelichting:* Dit is volgens de aanbevelingen over bestandsgrootte in Annex C van de
274 ISO 28500:2017 standaard.

5.5.3275 *Bestandsnamen*

276 *Eis:* De bestandsnamen voldoen aan het bestandsnaampatroon <ISIL-code van
277 archiefvormer>-<datum en tijd start harvest>-<volgnummer>-<domeinnaam van
278 seed>.warc.

279 *Toelichting:* Dit is volgens de aanbevelingen over bestandsnaamgeving in Annex C
280 van de ISO 28500:2017 standaard. Een voorbeeld. De eerste WARC van een
281 webarchief van de GahetNA website van het Nationaal Archief die op 31 oktober
282 2017 om 10:15:00 startte: NL-HaNA-20171031101500-001-www.gahetna.nl.warc.

5.5.4283 *Compressie*

284 *Eis:* De webarchiefbestanden zijn niet gecompriemd.

285 *Toelichting:* Dit voorkomt dat de webarchiefbestanden in de toekomst onleesbaar
286 zijn. De WARC-standaard staat compressie wel toe, maar deze richtlijn niet.

5.5.5287 *Digitaal waarmerk*

288 *Eis:* Van elk WARC-bestand is direct na creatie een checksum berekend die middels
289 een elektronische handtekening is ondertekend. Deze handtekening voldoet
290 tenminste aan een ETSI "BASELINE LT"-niveau, waarbij het
291 ondertekeningscertificaat, de certificaathierarchie, intrekingsinformatie en een
292 tijdstempel zijn opgenomen. Het ondertekeningscertificaat is uitgegeven door een
293 Trust Service Provider uit de PKI-overheid hierarchie.

294 Voorafgaand aan overbrenging naar een archiefbewaarplaats valideert de
295 archiefvormer alle over te brengen WARC-bestanden door de zowel de checksums
296 van de WARC-bestanden te controleren, als de handtekeningen van de checksums.

297 *Toelichting:* Met een digitaal waarmerk is op elk later tijdstip betrouwbaar vast te
298 stellen (onafhankelijk van waar en hoe de bestanden bewaard worden):

- 299 • Wanneer het bestand is gemaakt.
- 300 • Door wie het bestand is gemaakt.
- 301 • Dat het bestand na creatie niet is aangepast.

302 Dat vergroot het vertrouwen in de juistheid van de inhoud. Hoe vroeger in het
303 bewaarproces een checksum toegevoegd wordt, hoe beter de betrouwbaarheid van
304 de webarchiefbestanden kan worden gecontroleerd.

5.5.6305 *Volledigheid*

306 *Eis:* De webarchiefbestanden bevatten alle webpagina's met alle bijbehorende
307 onderdelen en bijlagen die binnen de harvestingscope vallen. Tenzij deze tijdens de
308 harvestingscope niet beschikbaar zijn. In dat geval is daarvan een melding opgenomen.

309 *Toelichting:* Al deze onderdelen zijn noodzakelijk om een website later volledig te
310 kunnen weergeven. Een uitzondering geldt voor onderdelen die om technische of
311 inhoudelijke redenen niet geharvest kunnen worden (opgenomen en toegelicht op
312 de blacklist).

5.5.7313 *Alle verzoeken en antwoorden*

314 *Eis:* De webarchiefbestanden bevatten ongewijzigd en integraal alle verzoeken
315 (requests) aan de webserver en alle antwoorden (responses) van de webserver die

CONCEPT

316 bij het harvesten zijn verzonden en ontvangen. Er is geen filtering toegepast en er
 317 zijn geen delen van de pagina herschreven. URL-rewriting is niet toegestaan.
 318 Foutmeldingen, omleningen en serverfouten worden ook allemaal opgenomen in het
 319 webarchiefbestand.

320 *Toelichting:* Om de authenticiteit en betrouwbaarheid van de informatie te
 321 beschermen is inzicht nodig in de automatische communicatie tussen de webserver
 322 en de harvestingdienst tijdens het proces van harvesten. Zo kan bijvoorbeeld
 323 geregistreerd worden dat een pagina niet mee-geharvest is omdat er op dat
 324 moment een serverfout was bij de archiefvormer.

5.5.8325 *Verplichte velden*

326 *Eis:* Naast de velden die verplicht zijn volgens de WARC-standaard, bevat elk
 327 webarchiefbestand ook ten minste de volgende velden (zoals beschreven in
 328 ISO28500:2017):

- 329 • In het Warcinfo-blok:
 - 330 ○ Software.
 - 331 ○ Format.
 - 332 ○ Robots.
- 333 • In de header van elk WARC-bestand:
 - 334 ○ WARC-Block-Digest.
 - 335 ○ Content-Type.
 - 336 ○ WARC-Warcinfo-ID.
- 337 • In de request- en response-blokken:
 - 338 ○ WARC-Target-URI.

339 Het is toegestaan andere velden te gebruiken dan die in de WARC-standaard staan.

340 *Toelichting:* Met deze extra velden kunnen de echtheid en onderlinge relaties van
 341 WARC-bestanden beter worden beschermd.

5.5.9342 *Leesbaar in gangbare viewer*

343 *Eis:* De webarchiefbestanden zijn leesbaar in een gangbare en leverancier-
 344 onafhankelijke viewer. Bij voorkeur open source. Voorbeelden van zulke viewers zijn
 345 The Wayback Machine en de Webarchiveplayer.

346 *Toelichting:* Een webarchiefbestand is pas bruikbaar als het ook gelezen kan worden
 347 met een viewer die voor iedereen beschikbaar is. Zonder daarbij afhankelijk te zijn
 348 van één specifieke leverancier.

5.6 349 **Metagegevens**

350 *Eis:* Leg voor de website als geheel en voor elk afzonderlijk webarchiefbestand
 351 metagegevens vast volgens de 'Richtlijn metagegevens Overheidsinformatie'. Deze
 352 metagegevens bestaan minimaal uit:

- 353 • Identificatiekenmerk [2]
- 354 • Naam [4]
- 355 • Omschrijving [6]
- 356 • Datum voor webarchiefbestand of periode voor de hele website [9.1]
- 357 • Taal [11]
- 358 • Actor [15]
- 359 • Vorm [19]
- 360 • Vertrouwelijkheid [17]
- 361 • Openbaarheid [18]

CONCEPT

- 362 • Integriteit [20]

363 De nummering verwijst naar het onderdeel in de 'Richtlijn metagegevens
364 Overheidsinformatie: Beschrijving van entiteiten en elementen. Versie 2.5'.

365 De overheidsorganisatie en de archiefbewaarplaats bepalen samen in welke vorm de
366 metagegevens moeten worden aangeleverd bij overbrenging.

367 *Toelichting:* Door in de metagegevens duidelijke contextinformatie mee te leveren,
368 kunnen de websites duurzaam toegankelijk blijven.

5.7 369 **Kwaliteitscontrole**

370 *Eis:* Laat een deskundige of onafhankelijke partij met een steekproef controleren of
371 de webarchiefbestanden voldoen aan de eisen. Herstel fouten als dat technisch nog
372 mogelijk is. Neem maatregelen om herhaling te voorkomen als dat mogelijk is.

373 *Toelichting:* Webarchiefbestanden die niet aan de eisen voldoen, kunnen voor een
374 deel of helemaal onbruikbaar zijn. Ze kunnen inhoudelijk onvolledig of onjuist zijn,
375 technisch niet te beheren of onleesbaar. Achteraf zijn fouten vaak moeilijk of niet
376 meer te herstellen. Daarom is het belangrijk fouten op tijd te zien.

5.8 377 **Beveiliging**

378 *Eis:* Bewaar de webarchiefbestanden en bijbehorende metagegevens in een
379 informatiesysteem dat zo beveiligd is dat verlies of veranderingen voorkomen
380 worden. Doe dat volgens de beveiligingsnorm NEN-ISO/IEC 27002 en de invulling
381 van deze norm zoals die voor de organisatie geldt. Zoals de Baseline
382 Informatiebeveiliging Rijksdienst (BIR) of de Baseline Informatiebeveiliging
383 Gemeenten (BIG).

384 *Toelichting:* Dit garandeert dat gearchiveerde websites precies hetzelfde zijn als op
385 het moment van harvesting als iemand ze in de toekomst bekijkt.

5.9 386 **Eigenaarschap**

387 *Eis:* De overheidsorganisatie is en blijft volledig eigenaar van de gearchiveerde
388 websites. Ook als werkzaamheden voor webarchivering worden uitbesteed. Zo kan
389 de overheidsorganisatie zelf beslissen hoe en aan wie ze toegang verleent tot de
390 webarchieven. De overheidsorganisatie kan altijd webarchiefbestanden opeisen die
391 door derden worden beheerd.

392 *Toelichting:* Dit voorkomt dat de overheid geen zeggenschap meer heeft over de
393 webarchieven.

5.10 394 **Toegang**

395 *Eis:* De geharveste websites zijn voor iedereen online toegankelijk. Met minimaal de
396 volgende functies voor alle gangbare webbrowsers:

- 397 • Zoeken en bekijken van de versie van elke datum van elke webpagina of bijlage
398 in de webarchiefbestanden. Vanaf de start van de harvesting tot nu.
399 • Full tekst-zoeken door alle webpagina's en bijlagen in de webarchiefbestanden.
400 • Bij het aanklikken van een link naar een pagina of bijlage buiten de
401 harvestingscope, volgt een duidelijke melding dat deze niet beschikbaar is in het
402 webarchief. Met de optie om de huidige versie te bekijken van de webpagina of

CONCEPT

- 403 bijlage. Bijvoorbeeld voor webpagina's op de blacklist of buiten de scope van de
404 website.
- 405 • Bij het aanroepen van een functionaliteit op een gearcheiverde webpagina
 - 406 waarvoor server-side scripting nodig is, een duidelijke melding dat de
 - 407 functionaliteit niet beschikbaar is in het archief. Bijvoorbeeld bij een zoekfunctie,
 - 408 contactformulier of een interactieve kaart.
 - 409 • Mogelijkheid om de metagegevens te bekijken van de website en van elke
 - 410 webpagina of bijlage.

411 Voor webpagina's en bijlagen op de lijst van beperkte openbaarheid geldt:

- 412 • Deze zijn niet online voor iedereen vindbaar en opvraagbaar. Waar technisch
- 413 mogelijk is dit alleen beperkt tot dat deel waarvoor de beperking geldt.
- 414 • Deze zijn opvraagbaar onder de voorwaarden die van toepassing zijn op de
- 415 wettelijke gronden die in de lijst zijn aangegeven.

416 *Toelichting:* Het doel van deze eisen is dat iedereen gearcheiverde websites kan
417 bekijken. Bij beperkt openbare informatie worden de belangen gerespecteerd.

5.11418 Blijvend bewaren

419 *Eis:* Gearcheiverde websites worden blijvend bewaard. Maak daarbij geen
420 onderscheid in de verschillende onderdelen van een website, of het moment waarop
421 een onderdeel is gearcheiverd.

422 Staat in de selectielijst van de organisatie niet duidelijk dat websites voor
423 publiekscommunicatie blijvend worden bewaard? Voeg dit dan alsnog toe bij de
424 eerstvolgende revisie van de selectielijst.

425 *Toelichting:* Websites voor publiekscommunicatie zijn een vorm van officiële
426 overheidscommunicatie. Ze zijn een belangrijke bron van informatie over het
427 handelen van de overheid. In het bijzonder over de relatie die de overheid
428 onderhoudt met burgers en bedrijven. Websites laten ook zien hoe de overheid zich
429 door de jaren heen presenteert aan de maatschappij. Ze hebben dus ook een
430 culturele waarde. Daarom worden ze blijvend bewaard.

431 Misschien zijn niet alle onderdelen van een website voor de lange termijn even
432 belangrijk. Maar omdat het lastig is om een goed onderbouwd onderscheid te
433 maken, worden alle onderdelen van een website blijvend bewaard.

5.12434 Overbrenging

435 *Eis:* Breng de webarchiefbestanden met bijbehorende metagegevens na uiterlijk
436 twintig jaar over naar een archiefbewaarplaats. Doe dat in de vorm van
437 webarchiefbestanden met bijbehorende metagegevens. De overheidsorganisatie en
438 de archiefbewaarplaats bepalen samen wanneer en hoe vaak.

439 Aanvullend op deze richtlijn gelden voor elke archiefbewaarplaats specifieke eisen
440 voor het overbrengen van digitaal archief. Neem daarvoor contact op met de
441 betreffende archiefbewaarplaats.

442 *Toelichting:* Volgens de Archiefwet moeten overheidsorganisaties archieven die
443 blijvend bewaard moeten worden, overbrengen naar een archiefbewaarplaats. Deze
444 is speciaal toegerust op het blijvend bewaren van archieven. Zo blijven de
445 webarchieven duurzaam toegankelijk.

CONCEPT

5.13446 Besluit beperking openbaarheid

447 *Eis:* Neem voor overbrenging een 'Besluit beperking openbaarheid' en publiceer dat
448 besluit. Doe dat op basis van de lijst met beperkingen aan de openbaarheid.

449 *Toelichting:* De archieven bij de archiefbewaarplaatsen zijn in principe openbaar.
450 Maar soms zijn er redenen om de openbaarheid tijdelijk te beperken. Zorgdragers
451 kunnen daarom bij de overbrenging beperkingen stellen aan de openbaarheid. Het
452 opstellen van het 'Besluit beperking openbaarheid' is onderdeel van het proces
453 overbrengen van overheidsarchieven.

CONCEPT

6 454 Bijlage: Voorbeeld stappenplan invoeren webarchivering

455 Dit voorbeeld stappenplan helpt een overheidsorganisatie bij het goed archiveren
456 van haar websites volgens de eisen uit de richtlijn. De stappen zijn bedoeld als een
457 voorbeeld van hoe de eisen uitgevoerd kunnen worden in de aangegeven volgorde.

6.1 458 **Initiatief**

459 *Actie:* Zet webarchivering op de agenda van de managers die direct of indirect
460 verantwoordelijk zijn voor de websites van de organisatie. Zorg ervoor dat zij het
461 belang van webarchivering inzien en vervolgens als opdrachtgevers het initiatief
462 nemen om webarchivering in te voeren en daarop toezien.

463 *Doel:* Bereidheid van het management om webarchivering in te voeren. Dit is
464 onmisbaar als er kosten gemaakt worden.

465 *Betrokkenen:* Opdrachtgevers websites, CIO, hoofd communicatie, hoofd
466 informatievoorziening, zorgdrager (of vertegenwoordiger daarvan),
467 archiefbeheerder.

6.2 468 **Alleen of samen**

469 *Actie:* Bepaal op welke schaal webarchivering wordt ingevoerd. Per website, per
470 organisatieonderdeel, voor de hele organisatie of samen met andere
471 overheidsorganisaties. Wanneer gekozen wordt voor samenwerking, doorloop dan
472 samen de volgende stappen.

473 Als er meerdere opdrachtgevers zijn, spreek dan af hoe ze samen sturing geven.
474 Bijvoorbeeld door het onder te brengen bij een bestaand programma, project of
475 overlegorgaan. Bij het gezamenlijk uitvoeren van webarchivering blijft elke
476 individuele organisatie verantwoordelijk voor het archiveren van de eigen websites.

477 *Doel:* Omdat het invoeren van webarchivering weinig afhankelijk is van de
478 specifieke eigenschappen van een website, kan samenwerken veel voordelen
479 hebben. Het inzetten van gedeelde kennis en samen inkopen van een
480 harvestingdienst en toegangsdienst kan bijvoorbeeld goedkoper zijn.

481 *Betrokkenen:* Contenteigenaar, contentbeheerder, platformbeheerder.

6.3 482 **Projectleider**

483 *Actie:* De betrokken managers wijzen een projectleider aan voor het invoeren van
484 webarchivering.

485 *Doel:* Iemand wordt beschikbaar gesteld als aanspreekpunt en voor het uit (laten)
486 voeren van de benodigde vervolgacties.

487 *Betrokkenen:* Opdrachtgevers webarchivering (zie bij 'initiatief').

6.4 488 **Informeren**

489 *Actie:* Bepaal wie een rol spelen bij webarchivering. Informeer deze betrokkenen
490 over het waarom, wat en hoe van webarchivering. Bijvoorbeeld met een presentatie
491 en discussie over de richtlijn webarchivering. Behandel daarbij vragen en maak
492 keuzes als dat nodig is.

CONCEPT

- 493 *Doel:* Betrokkenen zijn voorbereid op de mogelijke gevolgen van het archiveren.
 494 Bijvoorbeeld op werkzaamheden, applicaties, rolverdeling, financiën en contracten.
 495 Zo zorg het stappenplan niet voor verrassingen die de uitvoering kunnen vertragen.
- 496 *Betrokkenen:* Projectleider webarchivering, contenteigenaar, contentbeheerder,
 497 platformbeheerder, harvestingdienstverlener (als die al bekend is),
 498 archiefbeheerder, archiefbewaarpplaats.

6.5 499 **Rollen verdelen**

- 500 *Actie:* Bepaal wie welke rol op zich neemt bij webarchivering. Afhankelijk van hoe
 501 een overheidsorganisatie het archiveren regelt, kunnen de rollen anders verdeeld
 502 zijn.
- 503 *Doel:* Voor iedereen is duidelijk wat van hem verwacht wordt.
- 504 De zorgdrager is volgens de Archiefwet verantwoordelijk voor het archiveren van de
 505 websites. Hij moet er bijvoorbeeld voor zorgen dat er genoeg mensen zijn en dat
 506 duidelijk is wie waarvoor verantwoordelijk is. Maar ook dat er genoeg geld en
 507 hulpmiddelen zijn en dat er een selectielijst is met bewaartermijnen. Een zorgdrager
 508 is meestal niet rechtstreeks betrokken bij het archiveren. Maar hij regelt op een
 509 hoger niveau de algemene archieffunctie van een organisatie. Vaak heeft de
 510 zorgdrager zijn archieftaken overgedragen aan de CIO of het hoofd bedrijfsvoering.
- 511 Alle andere rollen kan de organisatie zelf bepalen. Een overheidsorganisatie hoeft
 512 het harvesten niet zelf uit te voeren. Ze mag dat ook laten doen door een andere
 513 overheidsorganisatie of een bedrijf. Maar de overheidsorganisatie blijft er wel zelf
 514 verantwoordelijk voor.
- 515 Een gebruikelijke rolverdeling:
- 516 • *Contenteigenaar:* Neemt het initiatief voor het archiveren en kijkt of dat goed
 517 gebeurt.
 - 518 • *Contentbeheerder:* Zorgt voor de inhoudelijke en procesmatige informatie over
 519 de website (zie: 'Overzicht websites').
 - 520 • *Functioneel beheerder van het webplatform:* Verstrekt de technische gegevens
 521 over de website. Hij kan daarnaast de harvestingdienst en toegangsdienst
 522 leveren of een externe leverancier daarvoor aansturen.
 - 523 • *Harvestingdienstverlener:* Zorgt voor de harvesting.
 - 524 • *Toegangsdienstverlener:* Zorgt ervoor dat iedereen toegang heeft tot de
 525 geharveste websites.
 - 526 • *Archiefbeheerder (ook wel medewerker DIV genoemd):* Zorgt voor de
 527 overbrenging van de webarchieven. Daarnaast kan de archiefbeheerder ook de
 528 harvesting en toegang verzorgen en de webarchiefbestanden bewaren tot het
 529 moment van overbrenging.
 - 530 • *Beheerder archiefbewaarpplaats:* Beheert de overgebrachte webarchieven en
 531 ontsluit deze voor het publiek. Daarnaast verzorgen archiefbewaarpplaatsen
 532 richtlijnen, best practices, voorlichting en advies over (web)archivering.

6.6 533 **Plan en budget**

- 534 *Actie:* Stel een plan van aanpak op. Beschrijf daarin minimaal doel, doelgroep,
 535 verwachte tijdsinvestering en budget. Als er nog geen budget is vastgesteld, is het
 536 belangrijk eerst duidelijk te krijgen hoeveel geld er beschikbaar is. Laat de
 537 opdrachtgevers het plan goedkeuren.
- 538 *Doel:* Planmatig werken en bewaken van het budget en de tijdsbesteding.

CONCEPT

539 *Betrokkenen:* Zorgdrager, contenteigenaar, contentbeheerder, platformbeheerder,
540 harvestingsdienstverlener (als die al bekend is), archiefbeheerder,
541 archiefbewaarpplaats.

6.7 542 **Overzicht websites**

543 *Actie:* Zoek uit van welke websites de overheidsorganisatie verantwoordelijk is voor
544 de inhoud. Er zijn (sub)websites waarvan de overheidsorganisatie wel
545 verantwoordelijk is voor de inhoud, maar niet voor het platform. Bijvoorbeeld de
546 webpagina's van Twitter met tweets van de overheid. Neem die (sub)websites ook
547 mee bij het inventariseren.

548 Maak afspraken over wie de inventarisatie hoe bijwerkt als er grote wijzigingen in
549 websites zijn. Bijvoorbeeld als er een nieuwe website bij komt, of als er een stopt.
550 Degene die verantwoordelijk is voor de webarchivering moet deze wijzigingen
551 verwerken.

552 *Doel:* De inventarisatie levert informatie op om te bepalen welke websites
553 gearchiveerd moeten worden en wie daarbij een rol speelt. De
554 harvestingsdienstverlener heeft die informatie nodig voor het archiveren. En de
555 contenteigenaar om de kwaliteit van de webarchieven te (laten) controleren
556 Tegelijkertijd helpt de beschrijving van de websites om die websites in de toekomst
557 te vinden en te interpreteren.

558 *Betrokkenen:* Contenteigenaar, contentbeheerder, platformbeheerder,
559 archiefbeheerder, archiefbewaarpplaats.

6.8 560 **Harvestingsdienst**

561 *Actie:* Maak afspraken over het uitvoeren van de harvestingsdienst van de websites (de
562 harvestingsdienst). Richt binnen de organisatie zelf een harvestingsdienst in of sluit
563 een contract met een externe leverancier. De harvestingsdienstverlener moet vooraf
564 aangeven welke eisen voor websites gelden om ze te kunnen harvesten. Let op dat
565 de harvestingsdienst niet ten koste gaat van de beschikbaarheid van de website voor
566 gebruikers.

567 Is de harvestingsdienstverlener een externe leverancier? Spreek dan ook af wat er
568 gebeurt als die stopt met zijn dienstverlening. Zo kan de opdrachtgever nog steeds
569 beschikken over de webarchieven die de externe leverancier heeft gemaakt.

570 Deze richtlijn bevat geen informatie over het inrichten van een eigen
571 harvestingsdienst. Raadpleeg daarvoor de documentatie en kennisplatformen van de
572 betreffende software.

573 *Doel:* De harvestingsdienstverlener maakt de webarchieven. Goede afspraken over
574 een duidelijke rolverdeling en over eisen waaraan de harvestingsdienst en de
575 websites moeten voldoen.

576 *Betrokkenen:* Zorgdrager, contenteigenaar, harvestingsdienstverlener.

6.9 577 **Toegangsdienst**

578 *Actie:* Maak afspraken over het beschikbaar stellen van de geharveste websites aan
579 het publiek (de toegangsdienst). Richt deze dienst in binnen de organisatie zelf of
580 sluit een contract met een externe leverancier. De toegangsdienst moet vooraf
581 aangeven welke eisen gelden voor de te harvesten websites om ze goed
582 toegankelijk te maken.

CONCEPT

- 583 *Doel:* Vaak levert dezelfde leverancier de toegangsdienst en de harvestingsdienst,
 584 maar dat hoeft niet. Goede afspraken maken duidelijk wat de rolverdeling is en aan
 585 welke eisen de toegangsdienst en de websites moet voldoen. Deze richtlijn bevat
 586 geen informatie over het inrichten van een eigen toegangsdienst. Raadpleeg
 587 daarvoor de documentatie en kennisplatformen van de betreffende software.
- 588 *Betrokkenen:* Zorgdrager, contenteigenaar, toegangsdienst.

6.10589 Bewaarplaats

- 590 *Actie:* Bepaal waar de webarchiefbestanden direct na harvestingsdienst worden
 591 tot het moment van overbrengen. Dit kan bijvoorbeeld in een voorziening waar de
 592 zorgdrager ook zijn andere archieven bewaart (vaak aangeduid als DMS, RMA of e-
 593 depot). Of bij de leverancier van de harvestings- of toegangsdienst. Controleer of
 594 deze bewaarplaats voldoet aan dezelfde beveiligingseisen die de zorgdrager stelt
 595 aan de bewaarplaats van zijn andere archieven. Neem daarvoor contact op met de
 596 archiefbeheerder, die kent deze normen.
- 597 *Doel:* Beheren van webarchiefbestanden volgens dezelfde beveiligingsnormen als de
 598 andere digitale archieven van de zorgdrager.
- 599 *Betrokkenen:* Harvestingsdienstverlener, toegangsdienstverlener, archiefbeheerder.

6.11600 Beperkingen openbaarheid

- 601 *Actie:* Maak afspraken over het beheren van de lijst met webpagina's en bijlagen
 602 waarvoor beperkte openbaarheid geldt:
- 603 • Wie houdt de lijst met beperkingen bij?
 - 604 • Waar en hoe kunnen verzoeken voor opname op de lijst worden ingediend?
 - 605 • Wie bepaalt hoe of het verzoek wordt gehonoreerd?
 - 606 • Hoe wordt de leverancier van de toegangsdienst geïnformeerd over
 607 wijzigingen in de lijst?
- 608 *Doel:* Voorkomen dat uit het webarchief informatie onrechtmatig wordt verstrekt.
- 609 *Betrokkenen:* Contentbeheerder, beveiligingsbeambte, privacybeambte, functionaris
 610 voor de gegevensbescherming (FG), leverancier toegangsdienst.

6.12611 Start harvestingsdienst

- 612 *Actie:* Meld alle websites uit de inventarisatie aan bij de harvestingsdienstverlener
 613 zodat harvestingsdienst start. Lever de geharveste websites aan de toegangsdienstverlener
 614 zodat ze toegankelijk zijn.
- 615 Informeer vooraf de functioneel beheerders van de websites over de start van de
 616 jaarlijkse harvestingsdienst van de hele website. En over het dagelijks harvesten van de
 617 wijzigingen. Zo kunnen zij als dat nodig is maatregelen nemen om harvestingsdienst
 618 mogelijk te maken zonder dat de website daar last van heeft.
- 619 *Doel:* Websites bewaren en bewaarde sites beschikbaar stellen. Voorkomen dat de
 620 websites slecht werken door het harvesten of dat harvestingsdienst niet goed lukt.
- 621 *Betrokkenen:* Contenteigenaar, contentbeheerder, platformbeheerder,
 622 harvestingsdienstverlener.

CONCEPT

6.13623 **Verwijzing op website**

624 *Actie:* Zet op de gearchiveerde websites verwijzingen naar het archief, bijvoorbeeld
625 bij de zoekfunctie.

626 *Doel:* Gebruikers van de website kunnen de gearchiveerde pagina's gemakkelijk
627 vinden.

628 *Betrokkenen:* Functioneel beheerders van websites.

6.14629 **Kwaliteitscontrole**

630 *Actie:* Laat de kwaliteit van webarchiefbestanden regelmatig controleren. Gebruik
631 hierbij de kwaliteitseisen in deze richtlijn. De leverancier moet onjuistheden meteen
632 verhelpen en dingen veranderen om herhaling te voorkomen. Als de
633 harvestingsdienst steeds niet voldoet aan de kwaliteitseisen, ga dan naar een nieuwe
634 harvestingsdienstverlener.

635 *Doel:* Zorgen dat de webarchiefbestanden voldoen aan de kwaliteitseisen. Zonder
636 dat daar achteraf dure reparaties voor nodig zijn of herstel zelfs onmogelijk is.

637 *Betrokkenen:* Contenteigenaar, harvestingsdienstverlener, archiefbeheerder.

6.15638 **Wijzigingen in websites**

639 *Actie:* Neem in de beheerprocedures van de websites op dat de
640 harvestingsdienstverlener op tijd ingelicht wordt als een website start, stopt of nooit
641 meer wordt bijgewerkt. En ook bij een grote verandering waardoor veel webpagina's
642 in één keer veranderen.

643 Stop de harvestingsdienst als de website stopt of de inhoud nooit meer wijzigt. Het kan
644 gebeuren dat een website aangemeld wordt voor harvestingsdienst terwijl hij al niet meer
645 bijgewerkt wordt. In dat geval is alleen een eenmalige volledige harvestingsdienst nodig.
646 Incrementele harvestingsdienst heeft dan geen meerwaarde meer.

647 *Doel:* De harvestingsdienstverlener is voorbereid op grote wijzigingen in de websites.
648 Voorkomen van onnodige harvestingsdienst.

649 *Betrokkenen:* Contenteigenaar, contentbeheerder, platformbeheerder,
650 harvestingsdienstverlener.

6.16651 **Werk selectielijst bij**

652 *Actie:* Pas de selectielijst aan als er niet duidelijk in staat dat websites voor
653 publiekscommunicatie blijvend bewaard worden. Doe dat op het moment dat de
654 selectielijst toch al wordt bijgewerkt.

655 Het is wettelijk verplicht om in een selectielijst vast te leggen hoe lang
656 overheidsinformatie bewaard moet worden. Maar het archiveren van websites is nog
657 niet overal gebruikelijk. Daarom zal in sommige selectielijsten nog onvoldoende
658 beschreven zijn wat de bewaartermijnen voor websites zijn. Bekijk samen met de
659 archiefbeheerder of dit het geval is.

660 *Doel:* Formele duidelijkheid over de bewaartermijn van websites.

661 *Betrokkenen:* Zorgdrager, archiefbeheerder.

CONCEPT

6.17662 **Overbrengen**

663 *Actie:* Breng de webarchiefbestanden van jaargangen waarvan de
664 overbrengingstermijn is verlopen over naar een archiefbewaarplaats.

665 De overbrengingstermijn is wettelijk maximaal twintig jaar. Maar een
666 archiefbewaarplaats kan toestaan de webarchiefbestanden al eerder over te
667 brengen.

668 Wat de archiefbewaarplaats is verschilt per overheidsorganisatie. Ook de manier van
669 overbrengen kan per archiefbewaarplaats verschillen. Maak daarom al bij het
670 inrichten van webarchivering afspraken met de archiefbewaarplaats over wanneer,
671 hoe vaak en hoe de bestanden worden overgebracht.

672 Neem contact op met de eigen archiefbeheerder. Die weet wat de
673 archiefbewaarplaats is en wat de afspraken zijn voor het overbrengen.

674 *Doel:* Zorgen voor beheer van blijvend te bewaren webarchieven. Zorgen dat het
675 beheer onafhankelijk is van de kennis en het voortbestaan van de archiefvormer.

676 *Betrokkenen:* Zorgdrager, archiefbeheerder, archiefbewaarplaats.

7 677 Bijlage: Begrippen

7.1 678 **Archiefbeheerapplicatie of archiefsysteem**

679 Een archiefbeheerapplicatie of archiefsysteem is een informatiesysteem voor het
680 archiveren van informatie.

681 Een informatiesysteem is het geheel van bestanden, procedures, apparaten en
682 hulpmiddelen, ingericht door een persoon, groep of organisatie voor het uitvoeren
683 van zijn of haar taken. Een archiefbeheerapplicatie omvat dus alles wat nodig is om
684 te kunnen archiveren.

685 De applicaties die gebruikt worden in een digitaal archiveringssysteem kennen
686 verschillende vormen. Zoals een DMS, RMA of een e-depot.

687 Archiveren kan in elk informatiesysteem, als het maar voldoet aan de eisen van
688 duurzame toegankelijkheid. Archiveren is maar één van de functies van een
689 informatiesysteem.

7.2 690 **Archiefbeheerder**

691 Degene die het archief van een organisatie beheert tot dit naar een
692 archiefbewaarpplaats gaat. Vaak heeft de archiefbeheerder een andere naam,
693 bijvoorbeeld recordmanager. Het kan ook een taak zijn van de afdeling
694 documentaire informatievoorziening (DIV).

7.3 695 **Archiefbewaarpplaats**

696 Een archiefbewaarpplaats heeft de taak om overheidsinformatie te beheren die voor
697 altijd bewaard moet blijven. De Archiefwet 1995 heeft de archiefbewaarpplaatsen
698 aangewezen.

699 Nadat een overheidsorganisatie de webarchiefbestanden heeft overgedragen, is de
700 archiefbewaarpplaats verantwoordelijk voor het beheren en toegankelijk maken van
701 de archiefbestanden. Inclusief de toegang tot de archieven.

702 De overheid heeft verschillende archiefbewaarpplaatsen:

- 703 • *Nationaal Archief*. Beheert archieven van de rijksoverheden die landelijk werken,
704 zoals ministeries en zelfstandige bestuursorganen (ZBO's). Bijvoorbeeld de
705 Kamers van Koophandel, Staatsbosbeheer en de Sociale Verzekeringsbank. Het
706 Nationaal Archief is ook het Regionaal Historisch Centrum voor de provincie Zuid-
707 Holland.
- 708 • *Regionaal Historisch Centrum (RHC)*. Hier worden de archieven van de
709 aangesloten provincies, gemeenten en waterschappen beheerd. En van regionaal
710 werkende rijksoverheden zoals rechtbanken. RHC's zijn
711 samenwerkingsverbanden tussen het Rijk, provincies, gemeenten en
712 waterschappen. De RHC's bevinden zich in de provinciehoofdsteden.
- 713 • *Gemeentelijke archiefbewaarpplaats*. Sommige gemeenten hebben een eigen
714 archiefbewaarpplaats. Andere gemeenten werken samen, bijvoorbeeld door deel
715 te nemen aan een RHC of een ander samenwerkingsverband.
- 716 • *Waterschapsarchiefbewaarpplaats*. Sommige waterschappen hebben een eigen
717 archiefbewaarpplaats, andere werken samen. Bijvoorbeeld door deelname aan een
718 RHC of een ander samenwerkingsverband.

CONCEPT

7.4 719 **Archiefvormer**

720 De persoon, groep personen of organisatie(onderdeel) waar gearchiveerde
721 informatie origineel van afkomstig is.

7.5 722 **Archiveren**

723 Archiveren betekent het duurzaam toegankelijk maken van informatie. Zo kunnen
724 belanghebbenden die informatie raadplegen zolang dat nodig is.

725 Overheidsorganisaties zijn bij wet verplicht te archiveren. Artikel 3 van de
726 Archiefwet zegt er het volgende over:

727 “De overheidsorganen zijn verplicht de onder hen berustende
728 archiefbescheiden in goede, geordende en toegankelijke staat te brengen en
729 te bewaren, alsmede zorg te dragen voor de vernietiging van de daarvoor in
730 aanmerking komende archiefbescheiden.”

731 Archiefbescheiden betekent alle informatie die een overheidsorgaan heeft gemaakt
732 of gekregen bij het uitvoeren van zijn taken. De vorm van die informatie maakt niet
733 uit. Overheidswebsites vallen daar dus ook onder.

734 “In goede, geordende en toegankelijke staat te brengen en te bewaren” betekent
735 duurzaam toegankelijk maken. Toegankelijk betekent vindbaar, beschikbaar,
736 leesbaar, interpreteerbaar en betrouwbaar. Duurzaam betekent dat de informatie
737 toegankelijk blijft voor zolang als nodig. Ook al verandert er van alles, bijvoorbeeld
738 informatiedragers, formaten, organisaties of gebruikerswensen.

739 Meer informatie over ‘duurzaam toegankelijk’ staat in de [DUTO-wiki](#) van het
740 Nationaal Archief.

7.6 741 **Beheerder archiefbewaarpplaats**

742 Organisatie of voorziening voor het blijvend bewaren van archieven. De overheid
743 heeft verschillende archiefbewaarpplaatsen zoals het Nationaal Archief, de Regionaal
744 Historische Centra of gemeentelijke archiefbewaarpplaatsen.

7.7 745 **Bijlage (bij een webpagina)**

746 Een bestand waarnaar een hyperlink verwijst vanuit de webpagina. Bijvoorbeeld een
747 pdf- of videobestand. Het is dus niet een andere webpagina.

7.8 748 **Checksum**

749 Een middel om te controleren of een digitaal bestand niet beschadigd is tijdens
750 bewaren of transport over een netwerk. Een checksum wordt bij de creatie van een
751 digitaal bestand of net voor het versturen over een netwerk berekend door een
752 algoritme. Het algoritme gebruikt daarvoor de bitsequentie van het digitale bestand.

753 De checksum kan op ieder willekeurig tijdstip in de toekomst, zoals na ontvangst
754 van het digitale bestand, opnieuw berekend worden met het algoritme.

755 Als tweede checksum overeenkomt met de eerste checksum, is vrijwel zeker dat het
756 bestand niet gewijzigd of beschadigd is geraakt tijdens het bewaren of versturen.

757 Er bestaan verschillende checksum algoritmes zoals MD5, SHA1, SHA256 of
758 SHA512.

CONCEPT

7.9 759 Compressie

760 Een techniek waarmee je de omvang van elektronische gegevens verkleint. De
761 kwaliteit van compressie hangt af van twee zaken: de toegepaste compressie-
762 algoritmen en de informatieobjecten waarop compressie wordt gedaan.

7.10 763 Contentbeheerder

764 Degene die de inhoud van een website maakt en onderhoudt. Dit kan de
765 contenteigenaar zijn of, of iemand die deze opdracht kreeg van de contenteigenaar.

7.11 766 Contenteigenaar

767 Degene die de inhoud van een website bepaalt. De contenteigenaar is de
768 zogenaamde archiefvormer.

7.12 769 Content Management Systeem (CMS)

770 Een programma of de software waarmee bijvoorbeeld een webredactie webpagina's
771 maakt en beheert.

7.13 772 Functioneel beheerder webplatform

773 Degene die het platform beheert waarmee een website wordt gemaakt en
774 onderhouden.

7.14 775 Harvesting

776 Zie [webarchivering](#), [client-side archivering](#).

7.15 777 Harvestingdienst

778 Een harvestingdienst omvat alle benodigde activiteiten voor het uitvoeren van de
779 harvesting op basis van de eisen van de opdrachtgever en met de afgesproken
780 frequentie.

7.16 781 Harvestingdienstverlener

782 Degene die de harvesting van een website verzorgt op basis van de lijst met
783 websites die de opdrachtgever heeft opgesteld. Daarvoor gelden de eisen in deze
784 richtlijn, eventueel aangevuld met eisen die de opdrachtgever zelf stelt. Belangrijk is
785 om dit alles duidelijk vast te leggen.

7.17 786 Harvestingfrequentie

787 De harvestingfrequentie geeft aan hoe vaak je een website archiveert. Er zijn
788 verschillende mogelijkheden: éénmalige harvesting, periodieke harvesting en real-
789 time harvesting.

7.17.1 790 Eenmalige harvesting

791 Dit gebruikt je voordat je een website verwijdert of vervangt.

CONCEPT

7.17.292 *Periodieke harvesting*

793 Bij periodieke volledige harvesting archiveer je steeds om de zoveel tijd. Dat
794 betekent wel dat veranderingen aan de site tussen twee momenten van harvesten
795 niet in het archief komen.

796 Een variant hierop is incrementele harvesting. Hierbij archiveer je alleen de
797 webpagina's die sinds de vorige harvest zijn gewijzigd of verwijderd.

7.17.298 *Real-time harvesting*

799 Bij real-time harvesting stuurt het CMS bij het toevoegen, wijzigen en verwijderen
800 van een webpagina een signaal naar de harvester. Die archiveert de pagina dan
801 (opnieuw).

7.18802 **Harvestingscope**

803 De afbakening van welke webpagina's en bijlagen een harvester moet binnenhalen.
804 Deze wordt meestal door de onderste parameters bepaald.

805 Daarvoor is het nodig om vooraf een complete lijst te geven van alle pagina's van
806 die website. Dat is meestal onmogelijk. Omdat dit een hele lange lijst kan zijn en
807 omdat die steeds verandert. Daarom gaat een harvester anders te werk.

7.18.808 *Startpagina (seed-URL)*

809 De harvester begint met het binnenhalen van de opgegeven startpagina. Vanuit de
810 startpagina haalt de harvester vervolgens alle webpagina's en bijlagen binnen die
811 via hyperlinks bereikbaar zijn.

812 De startpagina voor harvesting is meestal de hoofdpagina van de website (home
813 page). Bijvoorbeeld www.rijksoverheid.nl. Maar het kan ook een subdomein zijn,
814 zoals het Actorenregister: actorenregister.nationaalarchief.nl. Of een sub-website,
815 bijvoorbeeld www.overheidsorganisatie.nl/over-ons/wie-is-wie.

7.18.816 *Weespagina's*

817 Weespagina's zijn webpagina's die de harvester niet kan bereiken door vanaf de
818 startpagina een hyperlinkpad te volgen. Weespagina's horen wel bij de website. Als
819 bekend is wat de weespagina's zijn, zet deze dan op een lijst. Zo weet de harvester
820 dat hij die pagina's ook moet archiveren.

7.18.821 *Site-URL en secundaire site-URL's*

822 Welke webpagina's, die bereikbaar zijn vanaf de startpagina of een weespagina,
823 worden gearchiveerd wordt bepaald door de site-URL. Dat is de URL waarmee alle
824 webpagina's en bijlagen van een website mee beginnen. Zoals www.overheid.nl.
825 Soms kan het gebeuren dat een deel van een website met een andere URL begint.
826 Met name voor bijlagen die op een andere server zijn opgeslagen. Als dat het geval
827 is kunnen nog een aantal secundaire site-URL's worden opgegeven.

7.18.828 *Blacklist*

829 Dit is een lijst met pagina's die je om inhoudelijke of technische redenen wilt
830 uitsluiten van harvesting. De harvester slaat deze pagina's over.

831 Een voorbeeld van een technische reden om een pagina niet te harvesten is een
832 agenda op de site. Zo'n agenda kan tot in het oneindige doorlopen in het verleden
833 en in de toekomst. Daardoor weet de harvester niet waar hij moet stoppen.

CONCEPT

834 Bij het opstellen van de blacklist kan gebruik worden gemaakt van de inhoud van
835 het robots.txt-bestand. Dit bestand wordt vaak gebruikt om indexeerbots niet te
836 laten verdwalen op een website.

7.18.837 *Scope van webpagina-componenten*

838 Bij het harvesten moet je kiezen welke onderdelen van een webpagina in het archief
839 moeten komen. Dit kan alles zijn, of één of meer hostingdomeinen, bestandstypen of
840 multimediaformaten.

841 Een webpagina is meestal opgebouwd uit verschillende onderdelen: HTML, CSS,
842 JavaScript, afbeeldingen, multimedia en links naar bijlagen zoals pdf-documenten.
843 Het kan zijn dat sommige onderdelen op een andere server staan dan de site zelf.

7.18.844 *Hops en paddiepte*

845 Een website kan te groot zijn voor sommige archieven. Dan moet je kiezen tot
846 hoever je wilt gaan met archiveren. Dat kan bijvoorbeeld door het aantal hops en de
847 paddiepte te beperken.

- 848 • Het aantal hops is het kleinste aantal links dat je moet volgen om een pagina te
849 bereiken vanaf de startpagina.
- 850 • De paddiepte is het aantal URL-elementen dat in de URL achter de domeinnaam
851 staat. Een voorbeeld. De homepage 'www.rijksoverheid.nl' heeft een paddiepte
852 van 0. Maar 'www.rijksoverheid.nl/actueel/agenda' heeft een paddiepte van 2
853 ('actueel' en 'agenda').

854 Door het aantal hops en de paddiepte te beperken, komt het dieperliggende deel
855 van de website niet in het webarchief.

7.18.856 *Scope + 1*

857 Vaak hebben webpagina's een link naar pagina's buiten de website. Dat kan een
858 andere site zijn van de eigen overheidsorganisatie. Maar ook een website van een
859 andere organisatie. Bij 'scope + 1' haal je bij harvesten ook deze externe pagina's
860 binnen. Maar de hyperlinks op die externe pagina volg je dan niet. Vandaar +1.

7.19861 **Overbrengen**

862 Informatie overdragen aan een archiefbewaarplaats. Dat is verplicht volgens de
863 Archiefwet voor overheidsinformatie die blijvend bewaard moet worden. De
864 archiefbewaarplaats verzorgt het verdere beheer. Na het overbrengen is de
865 informatie openbaar en gratis in te zien. Behalve de informatie waarvoor een
866 beperking aan de openbaarheid gesteld is.

867 Overbrengen moet wettelijk uiterlijk twintig jaar nadat de bewaartermijn gaat lopen.
868 Die bewaartermijn staat in de selectielijst. De zorgdrager kan met de
869 archiefbewaarplaats afspreken om de informatie eerder over te brengen. Daardoor
870 is de zorgdrager korter verantwoordelijk en kan hij het beheer overlaten aan een
871 specialist.

872 Overbrengen is meer dan alleen het overdragen van de archiefbestanden. De
873 archiefbewaarplaats wordt ook eigenaar de van de informatie, met eventuele
874 afspraken die daarbij horen. Meestal draagt de beheerder van de bestanden ze over.
875 Maar de zorgdrager is verantwoordelijk voor de overdracht.

876 Meer informatie over overbrengen is te vinden in de [DUTO wiki](#) en op de [website](#)
877 [van het Nationaal Archief](#).

CONCEPT

7.20878 Selectielijst

879 In een selectielijst staat hoe lang welke informatie bewaard moet blijven of wanneer
880 die vernietigd moet worden. De overheidsorganisatie moet er ook in schrijven
881 waarom het die keuzes maakt.

882 In de Handreiking waardering en selectie (Belangen in Balans) staat wat een
883 selectielijst is en hoe hij gemaakt wordt. De Archiefwet stelt geen bewaartermijnen
884 vast voor webarchivering. De wet beschrijft alleen de procedure voor het opstellen
885 en vaststellen van een selectielijst. Elke overheidsorganisatie moet de
886 bewaartermijnen zelf bepalen door belangen af te wegen. Dat noemen we
887 waarderen. Archiefinstellingen waaraan zij hun blijvend te bewaren informatie
888 overbrengen helpen daarbij.

889 Informatie die volgens de selectielijst tijdelijk bewaard wordt, moet van de
890 Archiefwet na die tijd vernietigd worden. Informatie die blijvend bewaard moet
891 worden, moet de overheidsorganisatie overbrengen naar een archiefbewaarplaats.
892 Dat moet na uiterlijk twintig jaar.

7.21893 Toegangsdiens

894 Een toegangsdiens zorgt ervoor dat belanghebbenden toegang hebben tot de
895 geharveste websites. De toegangsdiens omvat alle activiteiten die nodig zijn voor
896 het kunnen vinden, gebruiken of terugzoeken van informatie. Bijvoorbeeld het
897 aanbieden van een viewer waarmee de geharveste website bekeken kan worden.
898 Zonder toegang is het nut van webarchivering klein. De toegangsdiens kan los
899 aangeboden worden, maar ook geïntegreerd met de harvestingsdiens.

7.22900 Toegangsdiensverlener

901 Een toegangsdiensverlener is de partij die het verlenen van toegang tot de
902 geharveste websites aanbiedt.

903 De toegangsdiens kan door verschillende partijen geleverd worden. Dit kan
904 bijvoorbeeld de archiefvormer zelf zijn, maar ook de harvestingleverancier of een
905 derde partij.

7.23906 URL (Uniform Resource Locator)

907 Het adres van een webpagina of bijlage. Een URL begint meestal met *http://*,
908 *https://* of *ftp://*.

7.24909 WARC-bestand

910 WARC betekent Web ARChive format. Het is een ISO-standaard voor het opslaan
911 van het resultaat van een harvestactie (ISO 28500:2017). Een WARC-bestand
912 bevat:

- 913 • Verzoeken die de harvester naar de webserver heeft gestuurd.
- 914 • Reacties die de webserver heeft teruggestuurd naar de harvester.
- 915 • Metagegevens over het harvestingsproces.

916 Elk WARC-bestand bestaat uit één of meer WARC-records. Een WARC-record bestaat
917 uit een header en een block.

CONCEPT

918 Het archief van één website kan zijn opgeslagen in meerdere WARC-bestanden.
 919 Bijvoorbeeld omdat een website te groot is voor één bestand. In de metagegevens
 920 in een WARC-bestand kan daarvoor een verwijzing staan naar de andere bestanden.
 921 Zo kun je een website die in meerdere bestanden is gearchiveerd toch goed
 922 bekijken.

7.25.923 **Webarchivering**

924 Webarchivering is het archiveren van websites: Het duurzaam toegankelijk maken
 925 van de webpagina's en bijlagen op een website.

926 Omdat webpagina's vaak veranderen en verdwijnen, gaat webarchivering niet
 927 vanzelf. Hiervoor moeten aanvullende maatregelen genomen worden. Er zijn
 928 verschillende manieren om dat aan te pakken. Client-side archivering is de meest
 929 gebruikte.

7.25.930 *Client-side archivering (harvesting)*

931 Bij deze methode benadert een harvester de website zoals een gebruiker en 'ziet' de
 932 website dus alleen aan de voorkant: de client-side. De harvester haalt webpagina's
 933 op bij de webserver. Inclusief afbeeldingen, vormgeving en downloadbare bijlagen
 934 (zoals documenten en filmpjes).

935 Een voordeel is dat het niet ingrijpt in de techniek van de website. Er wordt immers
 936 alleen contact gemaakt zoals een willekeurige bezoeker van de website dat ook
 937 doet. Deze manier van harvesting kent daarmee echter ook een aantal beperkingen:

- 938 • Geharveste webpagina's zijn opgevraagd door één (anonieme) gebruiker.
 939 Webpagina's waarvan de inhoud afhankelijk is van kenmerken van de gebruiker
 940 (zoals een ingevuld formulier of de inhoud van cookies) kunnen daarom op deze
 941 manier niet gearchiveerd worden. Maar vooral kunnen webpagina's waarvoor de
 942 gebruiker moet inloggen niet geharvest worden.
- 943 • Tijdens het raadplegen van een geharveste webpagina is de achterliggende
 944 webserver niet meer beschikbaar in de toestand van tijdens het harvesten. Als
 945 voor de weergave van een webpagina interactie met de webserver nodig is, is
 946 het resultaat daarom niet gegarandeerd hetzelfde als op het moment van
 947 harvesting. In veel gevallen zal het zelfs een foutmelding kunnen geven. Zoals
 948 bij streaming video en interactieve kaartjes. Zolang de weergave in zijn geheel
 949 binnen een webbrowser wordt afgehandeld, bestaat deze beperking niet.
- 950 • Websites die niet meer online zijn kunnen niet geharvest worden. Dit kan soms
 951 opgelost worden door de website tijdelijk weer online te zetten.

7.25.952 *Server-side archivering*

953 Bij deze methode haal je de inhoud van een website op bij het CMS. Of het CMS
 954 zorgt zelf voor het archiveren. Dit kan alleen als de beheerder van het CMS daaraan
 955 meewerkt. Hiervoor moet je met ieder CMS afzonderlijk afspraken maken. En
 956 misschien moet het CMS aangepast worden.

957 Het voordeel van deze aanpak is dat je alle informatie in het CMS kunt archiveren.
 958 Inclusief eventuele geschiedenisinformatie, wijzigingen aan teksten en andere
 959 informatie die wel in het CMS zit maar niet aan de voorkant (client-side) zichtbaar
 960 is.

961 Een nadeel van deze aanpak is dat die sterk afhankelijk is van de techniek van het
 962 CMS. Ook moet je de gearchiveerde inhoud van het CMS nog bewerken om die goed
 963 toegankelijk te maken.

CONCEPT

7.25.964 *Transactie-archivering*

- 965 Bij transactie-archivering archiveer je tijdens het gebruik van een website het
 966 uitwisselen van data tussen de webbrowser en de webserver. Daarmee archiveer je
 967 alle interactie tussen gebruikers en de website. Transactie-archivering wordt (nog)
 968 weinig toegepast.
- 969 Het voordeel van transactie-archivering is dat je precies alle informatie archiveert
 970 die de gebruikers kregen en verstuurden. Inclusief de inhoud die afhankelijk is van
 971 de gebruiker en de informatie die de gebruiker terugstuurt (bijvoorbeeld
 972 formulieren). Maar ook de interactie met de server voor het bepalen van het gedrag
 973 van een webpagina. Met transactie-archivering kun je het hele websitebezoek van
 974 een gebruiker terughalen.
- 975 Maar transactie-archivering kan ook een paar grote nadelen hebben. De privacy van
 976 de gebruikers van de website is namelijk in het geding. Daarnaast levert het ook
 977 veel gegevens en dat vraagt veel opslagruimte.
- 978 Een alternatief voor harvesting kan zijn webrecording. Hierbij neem je een sessie
 979 van één (anonieme) gebruiker op. Maar dat is onpraktisch omdat die gebruiker dan
 980 alle links handmatig moet aanklikken. En dit moet dan ook nog eens worden
 981 herhaald bij alle wijzigingen aan de website.

7.26982 **Webbrowser**

- 983 Een webbrowser is een computerprogramma om webpagina's te bekijken. Populaire
 984 browsers zijn Internet Explorer, Microsoft Edge, Mozilla Firefox, Google Chrome,
 985 Safari en Opera.
- 986 Dezelfde webpagina kan er anders uitzien bij verschillende webbrowsers of
 987 verschillende instellingen. Zelfs een andere versie van een browser kan een pagina
 988 anders laten zien. Ook het apparaat waarmee je een webpagina bekijkt kan verschil
 989 maken. Bijvoorbeeld een laptop, tablet of smartphone. Meestal zijn de verschillen
 990 klein. Maar er kunnen belangrijke verschillen zijn voor de bezoeker van de
 991 webpagina.

7.27992 **Webpagina**

- 993 Een webpagina is een document dat je met een webbrowser kunt bekijken. Een
 994 webpagina heeft een webadres (URL) en wordt beschikbaar gesteld door een
 995 webserver.
- 996 Kenmerkend voor een webpagina is dat deze vaak hyperlinks bevat naar bijlagen
 997 (verwijzingen in de vorm van een URL). Daarmee vormen alle webpagina's samen
 998 een 'web' van informatie waarin je gemakkelijk kunt rondkijken.
- 999 De webserver maakt een webpagina zichtbaar in de vorm van een of meer
 1000 bestanden. De basis daarvan is een bestand met (X)HTML-code. Daarnaast kunnen
 1001 er aanvullende bestanden zijn voor de verschillende onderdelen van de webpagina.
 1002 Zoals voor afbeeldingen, scriptcode en style sheets.
- 1003 Vaak staat een overheidswebpagina op een server van een overheidsorganisatie,
 1004 maar dat hoeft niet. Het kan ook een webserver zijn van een bedrijf of andere niet-
 1005 overheidsorganisatie. Dit verandert niets aan de wettelijke verantwoordelijkheid
 1006 voor het archiveren van de webpagina. Als een overheidsorganisatie
 1007 verantwoordelijk is voor de inhoud, is ze ook verantwoordelijk voor de archivering.
 1008 Zo bevat twitter.com/Rijksoverheid overheidsinformatie omdat de overheid tweets

CONCEPT

1009 heeft geplaatst. Daarmee valt (de inhoud van) twitter.com/Rijksoverheid onder de
1010 Archiefwet.

7.28011 Webservice en webclient

1012 Een webservice is een computer die via een netwerk webpagina's en bijlagen levert
1013 aan andere computers. Een webclient is een computer die de webpagina's ontvangt
1014 en laat zien aan degene die de webpagina heeft opgevraagd.

1015 Om een webpagina of bijlage op te vragen, stuurt de client een request
1016 (verzoek) naar de webservice. De server reageert daarop met een response
1017 (reactie). Die response kan de opgevraagde webpagina bevatten of een melding
1018 waarom de webpagina niet geleverd wordt. Daarnaast bevat de response
1019 metagegevens over het afhandelen van het request.

7.29020 Website

1021 Een website is een verzameling webpagina's en bijlagen die bij elkaar horen.

1022 Bij webarchivering is het belangrijk om precies te beschrijven welke webpagina's en
1023 bijlagen bij de website horen. Alleen die komen in het archief van die website.
1024 Meestal wordt de afbakening van een website bepaald door een site-URL.
1025 Bijvoorbeeld "https://nos.nl/". Alle webpagina's en bijlagen waarvan het webadres
1026 begint met de site-URL horen dan bij de website.

1027 Het komt voor dat het webadres van bijlagen van een website niet met de site-URL
1028 begint. Bijvoorbeeld omdat die op een speciale server staan. In dat geval zijn er
1029 aanvullende URL's nodig om vast te leggen welke bijlagen bij de site horen.

7.30030 Zorgdrager

1031 Degene die volgens de Archiefwet binnen een overheidsorganisatie bestuurlijk
1032 verantwoordelijk is voor het archiveren. De Archiefwet bepaalt per type organisatie
1033 wie de zorgdrager is. Meestal is dat de bestuurder of het dagelijks bestuur. Voor een
1034 ministerie is het bijvoorbeeld de minister en voor een gemeente het college van
1035 burgermeester en wethouders.

1036 In het dagelijks taalgebruik wordt het begrip zorgdrager ook vaak gebruikt voor de
1037 organisatie die zorgt voor het archiveren van bepaalde informatie. Maar strikt
1038 genomen klopt dat niet.

CONCEPT

8 1039 Bijlage: Motivatie van gemaakte keuzes

1040 Bij het schrijven van deze richtlijn maakte het Nationaal Archief verschillende
1041 keuzes. Hieronder motiveren we die.

8.1 1042 **Beperking tot publiekcommunicatie**

1043 Deze eerste versie van de richtlijn beperkt zich tot publiekscommunicatie uit
1044 praktische overwegingen. We hebben gekozen voor een afbakening met een relatief
1045 groot effect en weinig inspanning. Verdere uitbreiding van de scope betekent een
1046 langere ontwikkeltijd en een ingewikkelder richtlijn. Dat zou ten koste gaan van het
1047 snel invoeren van webarchivering.

1048 Webpagina's voor publiekscommunicatie zijn een belangrijk deel van het gezicht van
1049 de overheid. Misschien wel het belangrijkste deel. Burgers en bedrijven kunnen hier
1050 rechten aan ontlenuen en het is een onderdeel van ons cultureel erfgoed. Niet-
1051 openbare en persoonlijke webpagina's zijn met harvesting moeilijker te archiveren.
1052 Bovendien spelen hier al snel privacyoverwegingen die het archiefbeheer
1053 compliceren.

1054 In latere versies van de richtlijn kan de scope uitgebreid worden. Afhankelijk van de
1055 behoeftes.

8.2 1056 **Beperking tot harvesting**

1057 Deze eerste versie van de richtlijn beperkt zich tot harvesting als
1058 archiveringstechniek. Daarvoor kozen we omdat dit op dit moment de enige
1059 bewezen, eenvoudige en generieke techniek is. Bovendien heeft die geen impact op
1060 de websites en de achterliggende processen en applicaties. Voor harvesting is open
1061 source software beschikbaar en zijn er meerdere commerciële dienstverleners. Dat
1062 maakt het ook een betaalbare techniek.

1063 We onderkennen de beperkingen van harvesting. Maar daarvoor zijn op dit moment
1064 geen oplossingen voorhanden met vergelijkbare voordelen.

1065 In latere versies van de richtlijn kan de gebruikte techniek uitgebreid worden.
1066 Afhankelijk van de technische mogelijkheden.

8.3 1067 **Waarom vallen niet alle sociale media onder de richtlijn?**

1068 Alle uitingen van de overheid op sociale media vallen onder de Archiefwet. Als het
1069 belang voldoende is, dan moeten ze worden gearchiveerd. Hoe dat moet gebeuren
1070 is echter niet eenduidig. De technieken achter de sociale media verschillen sterk.
1071 Een aantal sociale media maken (ook) gebruik van webpagina's. Deze vallen wel
1072 onder deze richtlijn. Maar andere sociale media maken gebruik van een eigen app,
1073 zoals WhatsApp. Om deze te archiveren zijn andere technieken dan webharvesting
1074 nodig. Dit valt buiten de scope van deze eerste versie van de richtlijn.

1075 Daar komt bij dat veel uitingen op sociale media niet onder publiekscommunicatie
1076 vallen. Twitter bijvoorbeeld wel, maar dat is dan ook via webpagina's bereikbaar.

CONCEPT

8.4 1077 **Gebruik van WARC-files**

1078 De WARC-standaard is feitelijk wereldwijd de standaard voor de opslag van
1079 gearchiveerde webpagina's en bijlagen. Daarom hebben we gekozen voor het
1080 gebruik van de WARC-standaard. De standaard staat niet ter discussie en wordt
1081 breed ondersteund. Een bruikbaar alternatief is bij het schrijven van deze richtlijn
1082 niet in beeld gekomen.

8.5 1083 **Blijvend bewaren van gearchiveerde websites**

1084 De keuze om websites blijvend te bewaren is gebaseerd op het Generiek
1085 Waarderingsmodel Rijksoverheid 2.0 (GWR) en de Selectielijst gemeenten en
1086 intergemeentelijke organen 2017. Samen gelden deze als richtlijn voor het grootste
1087 groot deel van de overheid.

1088 Het GWR stelt dat "eindproducten van publieksvoorlichting" (waar websites onder
1089 vallen) blijvend bewaard moeten worden (categorie 5.1). De gemeentelijke
1090 selectielijst geeft aan dat "onlinecontent ten behoeve van externe communicatie"
1091 blijvend bewaard moet worden (nr. 20.1.2).

8.6 1092 **Alle webpagina's en bijlagen blijvend bewaren**

1093 Binnen één website worden alle webpagina's en bijlagen en alle versies daarvan
1094 blijvend bewaard. Ten eerste omdat verschillen in bewaartermijnen zorgt voor
1095 onvolledig gearchiveerde websites. Alsof je selectief pagina's uit een boek scheurt.
1096 Dit maakt het gebruik in de toekomst moeilijker. Daarnaast is één bewaartermijn
1097 voor alles een eenvoudige en duidelijke regel. Dat voorkomt misverstanden.

1098 Verschil in bewaartermijnen kan zorgen voor extra werk bij het selecteren en
1099 vernietigen. Het kan er zelfs toe leiden dat er helemaal niets bewaard wordt omdat
1100 de verschillende termijnen het ingewikkeld maken.

1101 De keuze om geen verschil te maken in bewaartermijnen sluit aan bij het Generiek
1102 Waarderingsmodel Rijksoverheid 2.0 (GWR) en de Selectielijst gemeenten en
1103 intergemeentelijke organen 2017. Deze maken geen onderscheid in
1104 bewaartermijnen binnen de categorieën "eindproducten van publieksvoorlichting" en
1105 "onlinecontent ten behoeve van externe communicatie".

1106 Er is een argument tegen het blijvend bewaren van alle webpagina's en bijlagen. De
1107 kosten voor het beheer wegen misschien niet op tegen de meerwaarde van het
1108 bewaren. Maar op dit moment zijn daar geen aanwijzingen voor. Mocht blijken dat
1109 de kosten hierdoor wel te hoog worden, dan heroverwegen we deze keuze.

1110 Het bewaren van alle versies van webpagina's stuit soms op een praktisch
1111 probleem. Webpagina's kunnen op een willekeurig tijdstip veranderd of verwijderd
1112 worden. Zelfs meerdere keren op één dag. Maar periodieke harvesting van websites
1113 gebeurt op vaste tijdstippen. Daarbij wordt alleen de laatste wijziging tussen twee
1114 harvests opgemerkt en vastgelegd. Daarom schrijft de richtlijn minimaal één keer
1115 per dag harvesten voor.

1116 Nog beter is real-time harvesting. Daarbij geeft het websitebeheerplatform elke
1117 wijziging in een website meteen door aan de harvestingdienst. Deze kan de
1118 wijziging dan meteen vastleggen. Bij het opstellen van deze richtlijn was real-time
1119 harvesting nog geen standaard praktijk en is daarom niet als eis opgenomen.

CONCEPT

8.7 1120 **Waarom verzorgt het Nationaal Archief de webarchivering niet?**

1121 Het behoort niet tot de wettelijke taken van het Nationaal Archief om archieven te
 1122 vormen. Volgens de Archiefwet is elke overheidsorganisaties daar zelf
 1123 verantwoordelijk voor. Wel worden blijvend te bewaren (web)archieven van het Rijk
 1124 op termijn overgedragen naar het Nationaal Archief. Daarna neemt het Nationaal
 1125 Archief het beheer op zich. Daarnaast is het Nationaal Archief een kenniscentrum
 1126 voor archivering en helpt daarmee overheidsorganisaties om hun (web)archivering
 1127 in te richten. Deze richtlijn komt daaruit voort.

8.8 1128 **Overheidswebsites worden toch al door de KB gearcheeerd?**

1129 De Koninklijke Bibliotheek (KB) archiveert inderdaad al overheidswebsites.
 1130 Daarnaast zijn er meer culturele en wetenschappelijke instellingen die websites
 1131 archiveren. Zoals het Nederlands Instituut voor Beeld en Geluid (NIBG). Het KB-
 1132 webarchief archiveert een selectie van Nederlandse websites vanaf 2007. Deze sites
 1133 worden één tot twee keer per jaar geharvest. Iedereen met een KB-jaarpas kan het
 1134 KB-webarchief in de leeszaal van de KB bekijken. NIBG archiveert websites van
 1135 omroepen.

1136 Waarom moeten overheidsorganisaties zelf hun websites archiveren als andere
 1137 instituten dat al doen? De reden hiervoor is dat al deze instellingen niet voor de
 1138 Archiefwet archiveren, maar voor het invullen van hun eigen doelstellingen. Ze
 1139 maken daarbij eigen afwegingen welke websites zij interessant en belangrijk vinden
 1140 om te archiveren. Met een zelfgekozen frequentie, selectie, bewaartermijn en
 1141 kwaliteitskenmerken. Deze keuzes zullen zelden overeenkomen met de keuzes die
 1142 voortvloeien uit de Archiefwet. Bovendien hoeven deze instellingen zich daarover
 1143 niet te verantwoorden volgens de Archiefwet.

1144 Dit alles maakt dat webarchivering door culturele en wetenschappelijke instellingen
 1145 niet in de plaats kan komen van de wettelijke webarchivering. Wel hebben deze
 1146 instellingen soms jarenlange kennis en ervaring met webarchivering die zij graag
 1147 delen met overheidsorganisaties.

8.9 1148 **Archiveren van webbrowsers buiten scope**

1149 Webpagina's worden geraadpleegd met veel verschillende webbrowsers en
 1150 apparaten. Die bovendien steeds veranderen. De weergave van webpagina's is niet
 1151 gestandaardiseerd en zal daarom voor elke combinatie van webbrowser en apparaat
 1152 anders zijn. Vaak zijn de verschillen klein. Maar relevante verschillen zijn niet
 1153 uitgesloten. Vooral niet tussen de weergave nu en in de verre toekomst.

1154 Daarmee is de principiële vraag aan de orde wat er eigenlijk gearcheeerd wordt bij
 1155 harvesting. Niet de eigenlijke webpagina, maar slechts de code om een webpagina
 1156 van te maken. Moet dan niet ook de techniek gearcheeerd worden waarmee uit de
 1157 code de webpagina gemaakt wordt?

1158 Dit is een fundamentele discussie waar nog geen uitsluitsel over is. Het lijkt een
 1159 oplossing om de webbrowsers zelf ook te archiveren. Maar dan moeten we ook de
 1160 apparaten bewaren waarop die browsers draaien (of emuleren) en deze operationeel
 1161 houden. Dat is praktisch onhaalbaar.

9 1162 Bijlage: Aanleiding en opdracht van deze richtlijn

9.1 1163 **Aanleiding**

1164 Overheidswebsites vallen onder de werking van de Archiefwet 1995. Dit betekent
1165 dat de overheid haar websites moet archiveren. Eind 2016 onderzocht de
1166 Erfgoedinspectie (EGI) in hoeverre dit daadwerkelijk gebeurt bij de centrale
1167 overheid. Conclusie is dat overheden hun websites en sociale media onvoldoende
1168 archiveren. Dit leidt tot verlies van informatie en is een (juridisch) risico. De EGI
1169 noemt als conclusies:

- 1170 • De organisaties van de centrale overheid archiveren hun websites, een
1171 enkele uitzondering daargelaten, niet of onvoldoende.
- 1172 • Hetzelfde geldt in nog sterkere mate voor officiële uitingen op sociale media.
- 1173 • Dit leidt tot verlies van informatie en is een risico in het geval van een
1174 juridisch geschil.
- 1175 • Voor de langere termijn dreigt digitale informatie met cultuurhistorisch,
1176 maatschappelijk en en/of wetenschappelijk belang te verdwijnen.
- 1177 • Het ontbreekt aan een beleidskader en aan specificaties voor het archiveren
1178 van overheidswebsites en uitingen op sociale media.
- 1179 • Mede daardoor hebben overheidsorganisaties onvoldoende handvatten voor
1180 het archiveren van hun websites en officiële uitingen op sociaalnetwerksites.
- 1181 • Er is onvoldoende kennis over webarchivering bij de archiefvormende
1182 organisaties.
- 1183 • De taakverdeling tussen Nationaal Archief, Koninklijke Bibliotheek en de
1184 Dienst Publiek en Communicatie is, voor wat betreft webarchivering, bij de
1185 archiefvormende organisaties onvoldoende duidelijk.

1186 Het rapport leidde tot Kamervragen en daarmee tot een Kamerbrief waarin de
1187 minister van OCW het volgende toezegt:

- 1188 • Ik zal de Algemene rijksarchivaris [algemeen directeur van het Nationaal
1189 Archief] vragen een handreiking op te stellen en in beheer te nemen gericht
1190 op de archivering van websites en uitingen op sociale media.
- 1191 • In samenwerking met CIO-Rijk zal hij tevens een voorstel doen ten behoeve
1192 van implementatie van een archiveringspraktijk voor deze media bij
1193 organisaties binnen de centrale overheid.

9.2 1194 **Uitwerking**

1195 In deze richtlijn werkt het Nationaal Archief de eerste toezegging van de minister uit
1196 voor de hele overheid. Het plan voor de implementatie van de richtlijn voor de
1197 centrale overheid is een afzonderlijk document. Het is opgesteld in opdracht van de
1198 CIO-Rijk in samenwerking met het Nationaal Archief.

1199 De definitie 'webarchivering is het archiveren van websites' in deze richtlijn wijkt af
1200 van de definitie in het EGI-rapport ('webarchivering is een containerbegrip voor het
1201 binnenhalen en bewaren van websites en van uitingen op sociale media.') Het
1202 'binnenhalen en bewaren van websites' is slecht één techniek van archiveren. Het
1203 omvat bovendien niet alle aspecten van archiveren, zoals de vindbaarheid en
1204 beschikbaarheid. Verder omvat in onze definitie webarchivering alleen sociale media
1205 voor zover die via een website beschikbaar zijn. Omdat andere vormen van sociale
1206 media technisch heel anders werken dan websites.

CONCEPT

9.3 1207 **Opdracht**

1208 Het Nationaal Archief ontwikkelt en beheert de richtlijn in opdracht van de minister
1209 van OCW. De richtlijn wordt als kennisproduct onderdeel van de zakelijke
1210 dienstverlening van het Nationaal Archief. Sturing op de ontwikkeling en het beheer
1211 van de richtlijn vindt plaats via de reguliere governancestructuren. Onderdeel
1212 daarvan is dat deze richtlijn uiteindelijk vastgesteld zal worden door de
1213 Standaardisatieraad van het Nationaal Archief.

9.4 1214 **Verdere ontwikkeling**

1215 Het Nationaal Archief ontwikkelt de richtlijn in stappen. Het uitgangspunt is dat we
1216 bestaande kennis zo snel mogelijk beschikbaar stellen. Daarna kunnen verschillende
1217 archiefvormers die direct toetsen in de praktijk. Het Nationaal Archief verwerkt deze
1218 praktijkervaringen in de richtlijn.

1219 De stappen die deze richtlijn doorloopt:

- 1220 • Eerste conceptversie (december 2017).
- 1221 • Openbare reviewversie (tweede kwartaal 2018).
- 1222 • Vastgestelde eerste versie (na openbare review 2018)
- 1223 • Aanpassingen op basis van praktijkervaringen (verwacht vierde kwartaal
1224 2018).
- 1225 • Verdere uitbreiding van de scope (vanaf 2019).

1226 De planning is afhankelijk van de reacties op de verschillende versies van de
1227 richtlijn.